Volume 13 Number 11 1985 Nucleic Acids Research

Multiple sequence-specific DNA binding activities are eluted from chicken nuclei at low ionic strengths

M.A.Plumb, R.H.Nicolas, C.A.Wright and G.H.Goodwin

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

Received 12 February 1985; Accepted 2 May 1985

ABSTRACT

DNA sequence-specific binding proteins eluted from chicken erythrocyte and thymus nuclei, and fractionated as described by Emerson and Felsenfeld (19), have been investigated by filter binding and footprint analyses. The erythrocyte nuclear protein fraction specifically binds to at least two sites within the 5' flanking chromatin hypersensitive site of the chicken β^A-globin gene, and to a site 5' to the human β-globin gene. The major chicken β^A globin gene binding site ($(G)_{18}CGGGTGG$) and the human β-globin gene binding site ($(TA)_6(T)_8C(T)_4$) occur at or near sequences which are hypersensitive to S1 nuclease cleavage in supercoiled plasmids. Downstream, the second chicken β^A-globin gene binding site includes the β-globin gene CACCC consensus sequence. Filter binding studies also show other sequence specific binding activities to human N-ras and human (but not chicken) c-myc gene sequences.

INTRODUCTION

The chicken β-globin genes (ρ, β^H, β^A and ε) are a clustered multigene family (1,2) whose expression is developmentally regulated (3,4). Although the four chicken β-globin genes reside in an 8 Kb domain and are transcribed as independent transcription units from the same DNA strand, the ρ and ε genes are specifically expressed in embryonic (primitive red cells) erythrocytes whereas the β^H and β^A are specific for adult (definitive red cells) erythrocytes (3,4). In both adult and embryonic erythrocyte chromatin, the β-globin domain and several Kb of flanking sequences exist in a conformation which is highly sensitive to endonuclease cleavage compared to bulk chromatin (5-7), but is comparatively insensitive to DNAse I in presumptive erythroid precursor cells (7). Within this domain, sites in the 5' and 3' flanking regions of transcriptionally competent genes are found which are hypersensitive to double-stranded endonuclease cleavage (6-9). Two main types of DNAse I hypersensitive sites in chicken erythrocyte β-globin chromatin have been described (6-9): (i) hypersensitive sites which are developmentally regulated and (ii) sites which are present in both adult and embryonic erythrocyte chromatin.

The partial purification of sequence-specific DNA-binding proteins which bind to the 5' flanking gene regulatory sequences has now been described for a variety of systems including the Drosophila histone genes (10), the Drosophila heat shock genes (11-13), the steroid hormone induced genes associated with the mouse mammary tumour virus (14), polyoma virus (15), SV40 (16), Xenopus 5S genes (17), the human c-myc gene (18), the chicken β^A-globin gene (19), the chicken ovalbumin gene (20,39), and the chicken lysozyme gene (40).

The specific binding of a chicken erythrocyte nuclear protein fraction to the 5' flanking region of the chicken β^A-globin gene is especially interesting as Emerson and Felsenfeld (19) were able to show that the protein conferred nuclease hypersensitivity to chicken β^A-globin gene chromatin assembled in vitro. The nuclear endonuclease hypersensitive site in chicken β^A-globin chromatin is developmentally regulated (6-9) and has been mapped to a 200 bp region 60 bp to 260 bp 5' to the mRNA CAP site (8). In cloned supercoiled plasmids, two sequences in this region of the β^A-globin gene are especially sensitive to S1-endonuclease cleavage: the first is at an inverted repeat sequence 20 bp 5' to the ATAA consensus sequence (21), and the second occurs in a tract of 18 consecutive deoxyguanosine residues 183-201 bp 5' to the mRNA CAP site (9,22,23) indicating that an altered DNA secondary structure is inherent in this hypersensitive domain. This is supported by the observation that when core histones are reconstituted onto supercoiled β^A-globin DNA by gradient dialysis, this region retains its S1-nuclease sensitivity (23) suggesting that core nucleosomes bind poorly to that sequence. However, when chromatin is assembled in the Xenopus oocyte cell free nucleosome assembly system, the DNAseI hypersensitivity of the region is reconstituted only if the supercoiled plasmid DNA is preincubated with the erythrocyte nuclear protein fraction (19).

As there are other developmentally regulated hypersensitive sites in the β-globin chromatin domain (6-9), we have examined the β^A-globin gene binding chicken erythrocyte nuclear protein fraction for additional sequence specific binding activities. Using a filter binding assay and/or footprint analysis to detect sequence-specific binding the following questions have been approached: (i) does the protein fraction exhibit specific binding to other chicken β-globin sequences which may be involved in other developmentally regulated DNAse I hypersensitive sites.; (ii) is specific binding to non-globin genomic chicken DNA sequences detectable.; (iii) is binding to human genomic DNA sequences detectable and, if so, can they be related to the chicken genomic sequences.; and (iv) are one or more of these activities tissue specific.

Although it was not possible to detect specific binding to 25 Kb of cloned chicken genomic DNA sequences other than the 5' flanking sequences of the β^A-globin gene, we demonstrate that the chicken erythrocyte nuclear protein fraction will specifically bind at least three distinct genomic human gene sequences. The specific binding to the 5' flanking sequence of the human β-globin gene was examined in depth as binding also occurs at an S1 nuclease sensitive site in supercoiled human β-globin DNA (24) and occurs under comparable conditions to the binding to the chicken β^A-globin DNA despite the fact that a $(TA)_6(T)\overset{*}{8}C(T)_3$ sequence is bound on the former, whereas a $(G)_{18}C(G)_3T(G)_2$ sequence is bound in the latter. These observations, with the detection of the binding activities in non-erythroid thymus nuclei, are discussed in relation to the possible binding to altered secondary DNA structures.

MATERIALS

Adenosine 5'-[γ-^{32}P]triphosphate (5000 Ci/mmol) and cytosine 5'-[α-^{32}P] triphosphate (~400 Ci/mmol) were purchased from Amersham International, PLC, Bucks. The DNA sequencing system (Maxam-Gilbert procedure (32)) was from New England Nuclear. Restriction and DNA modifying enzymes were from BRL. Protease K was from Boehringer-Mannheim; calf thymus DNA-cellulose (5.7 mg DNA/g cellulose) was from Sigma; X-ray film was from Kodak Ltd (XAR5) or Fuji (RX). Deoxyribonuclease I (DNAse I) was from Millipore Corporation.

Recombinant DNA

Recombinant clones of the chicken β-globin domain were obtained either by subcloning EcoRI, EcoRI-HindIII or HindIII restriction fragments of the $\lambda C\beta G_2$ recombinant (26), or by subcloning the EcoRI-HindIII fragment of pBIBR15 (33) into pUC9 (25) ($\lambda C\beta G_2$ and pBIBR15 were kindly provided by D. Engels). The human β-globin gene clone pBF5 (27) was from A. Malcolm, but the BamH1-BamH1 insert was subsequently subcloned into pUC9 (pBF5.6). The chicken genomic c-*myc* gene clone (pCc-*myc*) was from K. Bister (28), the chicken histone gene cluster clone (pCH3.3e) (29) was from J. Welles and the human genomic N-*ras* gene subclone (pAT7.0) (30) was from A. Hall.

End-labelling restriction fragments

(i) DNA restriction fragments and [α-^{32}P]-dCTP (1.3-2 μCi dCTP/μg DNA) were incubated in the presence of the large fragment of *E. coli* polymerase 1 (Klenow fragment, 1u/μg of DNA) in T4 polymerase buffer (31) for 30 minutes at room temperature. The reaction was stopped by the addition of 200 μl of a solution containing 0.4M potassium acetate, 10 mM tris-HCl (pH 8.0), 1 mM EDTA

and 50 µg/ml yeast tRNA. The solution was extracted twice with 2 volumes of buffered phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v), once with 1 vol. of chloroform:isoamyl alcohol (24:1, v/v) and precipitated with 3 vol. of ethanol at -70°C. The nucleic acid pellet was reprecipitated from 100 µl of a 0.1M potassium acetate, 10 mM tris-Cl (pH 8.0), 1 mM EDTA solution and then dissolved in 10 mM tris (pH 8.0), 1 mM EDTA.

(ii) Blunt-end or 5' overhang restriction sites were dephosphorylated by bovine alkaline phosphatase (BAP, ∿ 1u/50 µg of DNA) in a solution containing 10 mM tris-Cl (pH 8.0), 80 mM NaCl, 0.1 mM EDTA at 37°C for 15 minutes. 3'-overhand restriction sites were first treated with T4 DNA polymerase (1 u/µg DNA) in T4 polymerase buffer (31), containing 0.2 mM of each of the four deoxyribonucleotides, for 5 min. at 37°C. After deproteinization by organic extractions and two ethanol precipitations, the DNA was treated with BAP as described above.

Phosphatased DNA was end labelled with [γ-^{32}P]-ATP at 37°C for 30 min. in kinase buffer (31) containing 10 mM spermidine and T4 polynucleotide kinase (2u/µg of DNA). The DNA was deproteinized as described above, restricted with a second restriction enzyme, and the 5'-end labelled restriction fragments resolved by 1.5% (w/v) low melting point agarose (BRL) gel electrophoresis. DNA was purified from the molten agarose by chromatography through an Elutip-D (Schleicher and Schuell) mini column in 5 vol. of low salt buffer (0.2M NaCl, 20 mM tris-Cl (pH 7.4) 1 mM EDTA) at 42°C. The column was washed with 5 ml of low salt buffer, eluted with 600 µl of high salt buffer (1.0M NaCl, 20 mM tris-Cl (pH 7.4), 1 mM EDTA), and the DNA concentrated by ethanol precipitation.

Chicken nuclear protein fraction preparation

Nuclei from 15 day or adult chicken erythrocytes were prepared and extracted with 0.15M NaCl, 10 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$, 0.5 mM PMSF as described by Emerson and Felsenfeld (19). Nuclei were then extracted with 0.3M NaCl, 10 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$, 0.5 mM PMSF. The 0.15M and 0.3M NaCl nuclear extracts were either pooled or processed separately as described (19). Nuclei from 3-week old chicken thymus were prepared by homogenization in ice cold solution A (10 mM tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.25M sucrose, 0.2% (w/v) triton X-100, 0.5 mM PMSF), filtered through 3 layers of gauze and nuclei pelleted at 1600g for 30 minutes. The nuclei were washed twice with solution A, resuspended in 10 mM tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.5 mM PMSF and slowly made either 0.15M or 0.3M with respect to NaCl by the dropwise addition of a 4M NaCl solution. The suspension was incubated at 37°C for

10 min., the supernatant recovered by centrifugation at 2000g for 10 min., and fractionated by ammonium sulphate precipitation (0.35g/ml) and calf thymus DNA-cellulose affinity chromatography as described (19). DNA-binding proteins were sequentially eluted from the DNA cellulose with 100 mM (C3 fraction), 250 mM (C4 fraction) and 500 mM (C5 fraction) ammonium sulphate. Protein concentrations were estimated by densitometry of PAGE-blue 83 (BDH) stained proteins resolved by SDS polyacrylamide gel electrophoresis.

Filter binding assay

Up to 1 µg of protein were bound to ~ 150 ng of Klenow α-^{32}P-end labelled restriction fragments for 90 min. at 4°C in binding buffer (0.1M NaCl, 50 mM tris-Cl (pH 8.0), 3 mM $MgCl_2$) containing up to 20 µg of competitor E. coli DNA in a final volume of 30 µl. The solution was filtered through a nitrocellulose filter (25 mm diameter, Scheicher and Schuell BA85) and the retained fragments eluted and concentrated for 1.5-2% agarose gel electrophoresis and autoradiography essentially as described (19).

Footprint analysis

T4 polynucleotide kinase end-labelled DNA restriction fragments (10-40 ng, 0.5-1.5x10^6 cpm/µg) were incubated with protein (0-25 µg) in a final volume of 30 µl (made up with binding buffer) or a final volume of 100 µl (made up with storage buffer (19)), in the presence or absence of 500 ng of E. coli competitor DNA, for 90 min. at 4°C. The solution was made 12 µg/ml with respect to DNAse I, incubated at room temperature for 30 sec., and digestion halted by the addition of 100 µl of stop buffer (100 mM tris-Cl (pH 8.0), 10 mM EDTA, 0.5% (w/v) SDS, 70 µg/ml protease K, 150 mM NaCl, 0.1 mg/ml tRNA). After a 30 min. incubation at 37°C, the solution was deproteinized by organic extractions, ethanol precipitated, rinsed in 95% ethanol and redissolved in 95% (w/v) formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol for denaturing 6% (w/v) polyacrylamide, 50% (w/v) urea gel electrophoresis. The same end-labelled fragment was sequenced in parallel by the chemical modification procedure described by Maxam and Gilbert (32).

RESULTS

Sequence-specific binding to the chicken β-globin domain

Nuclei prepared from 15-day old embryonic or from mature erythrocytes were extracted first with 0.15M NaCl and then with 0.3M NaCl, and proteins fractionated by ammonium sulphate precipitation and DNA cellulose affinity chromatography as described in Methods. Specific binding to the chicken β^A-globin gene was analysed by incubating the protein extracts with

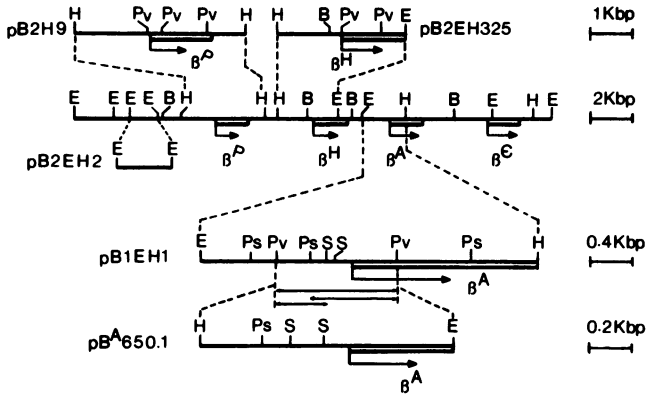


Figure 1:
The chicken β-globin domain.
Restriction map of the β-globin domain (2,3) showing fragments which were subcloned into pUC9. E: EcoRI; H: HindIII; B: BamHl; Pv: Pvull; Ps: Pstl; S: Smal. pβ^{A}650.1 was isolated by blunt end ligation of the internal 575 bp Pvull fragment of pBIEHl into the HincII site of pUC9. (⟶) denotes the direction of transcription (2,3), and (⟷) represents the pBIEHl restriction fragments which are specifically retained by the erythrocyte C4 fraction in filter binding assays.

end-labelled restriction fragments of the plasmid pBIEHl (see Fig.1) in the presence of increasing amounts of E. coli competitor DNA, and then assaying for the specific retention of a fragment after filtration through a nitrocellulose filter.

As shown in Fig.2, C4 protein fractions (eluted from DNA-cellulose with 250 mM $(NH_4)_2$ SO_4) from adult erythrocyte nuclear 0.15M and 0.3M NaCl extracts selectively bind the 575 bp Pvull fragment from pBIEHl. Approximately 1.5-2 times more binding activity is extracted in the 0.3-C4 fraction than in the 0.15-C4 fraction (Fig.2b), indicating that two populations can be differentially eluted from nuclei. Additional filter binding assays with different restriction digests of pBIEHl confirmed Emerson and Felsenfeld's (19) localization of the binding site to a Pstl-Smal fragment (position -20 to -109 relative to the CAP site, see Fig.1 and Fig.3). The chicken β^{A}-globin gene binding activity can also be extracted with embryonic 15 day old chick erythrocyte nuclei by both 0.15M and 0.3M NaCl, and filter binding analyses of serial dilutions of the C4 fractions suggest that the binding activity is more abundant (∿ 5 fold) in embryonic than in mature erythrocyte nuclei (data not shown). It should also be noted that preferential retention of the vector DNA is often observed, particularly when the DNA fragments are above 2 Kb in

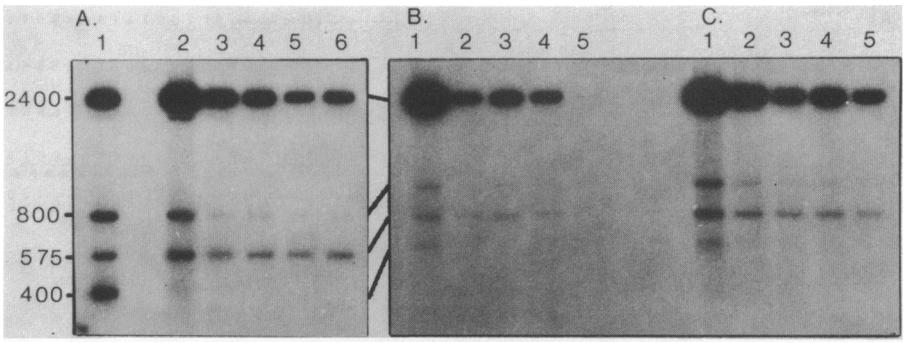


Figure 2:
Filter binding analysis of the erythrocyte C4 fraction to the chicken β^A-globin gene.
An end-labelled EcoRI, HindIII and PvuII triple digest of pBIEH1 (150ng) was bound to C4 fractions prepared from consecutive 0.14M and 0.3M NaCl nuclear extracts from mature chicken erythrocytes in the presence of increasing amounts of competitor E.coli DNA, and filtered as described in Methods.
(A) Lane 1: Control restricted DNA (75 ng). Lanes 2-6: ~300ng of the 0.3M NaCl C4 proteins fraction bound in the presence of 0,1,2,4 and 8µg of competitor DNA respectively.
(B) Lanes 1-4: ~450ng of the 0.15 M NaCl C4 protein fraction bound in the presence of 0,1,2 and 4µg of competitor DNA respectively. Lane 5: Control filter binding in the absence of protein.
(C) Lane 1-5: ~900ng of the 0.15 M NaCl C4 protein fraction bound in the presence of 0,1,2,4 and 8µg of competitor DNA respectively.
Similar results were obtained with ~180ng or ~65ng of C4 protein fractions prepared from 0.15 M or 0.3 M NaCl nuclear extracts from 15 day old chicken erythrocytes, and, with ~250ng of a C4 fraction prepared from pooled 0.15 M and 0.3 M NaCl extracts from mature erythrocytes.

length (also see (19)).

The precise binding site was mapped by footprint analysis using a pUC9 subclone of the pBIEH1 plasmid which contains the 575 bp PvuII fragment ($pB^A650.1$, see Fig.1). $pB^A650.1$ was 5' end-labelled at either the HindIII (Fig.4) or PstI (Fig.5) sites, the insert isolated after secondary digestion with EcoRI, and incubated in the presence of increasing protein with or without competitor DNA. As can be seen in Figs.4 and 5, at least two distinct sites are protected from DNAse I cleavage, and these were located using sequencing reactions run in parallel (Fig.5). One site, $(G)_{18}CGGGTGG$, extends from position -201 to -177 bp relative to the CAP site (see Fig.3a, site A) and the second commences 15 bp 3' to the G-rich site (position -161) and at an intermediate protein concentration (~ 2.5-5 µg protein/10 ng DNA) extends to position -115 (see Fig.3a, site B+B'). However, if the protein concentration is reduced to 1 µg/10 ng DNA, the second footprint decreases to position -161

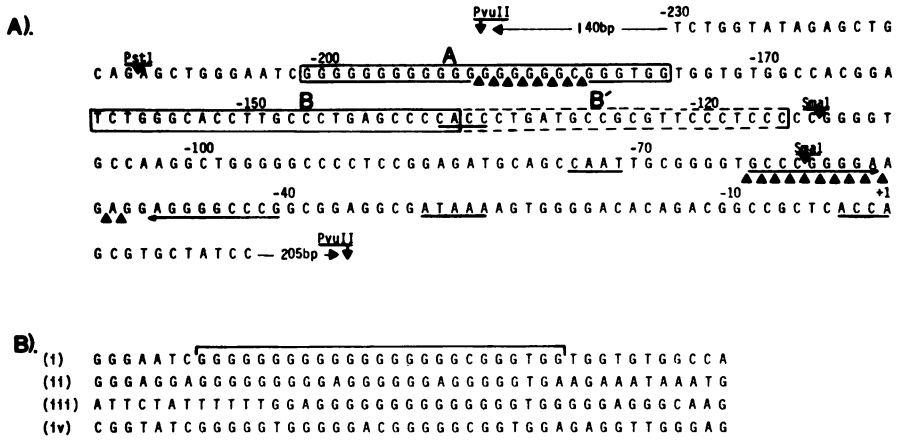


Figure 3A:
Partial sequence of the pβ^{A}650.1 insert (Fig.1 and (33)). Sequences which are protected from DNAse I cleavage in footprint analyses are boxed (Figs. 4 and 5). The arrowheads denote S1 cleavage sites in supercoiled plasmids as reported (21-23), and (◄— —►) denotes the inverted repeat. Sequences underlined are published 5' consensus sequences (33,44).
Figure 3B:
Comparison of the $(G)_{18}$CGGGTGG chicken β^A-globin gene sequence (i), to similar sequences in (ii) the 5' region of chicken c-myc gene (35), (iii) the first intron of chicken c-myc gene (35), and (iv) the 5' region of the human c-myc gene (34). Only the chicken β^A-globin sequence (i) is specifically bound by the erythrocyte C4 protein in a filter binding assay.

to -137 (Fig.4 and 3a, site B), but merges with the G-rich site if the protein concentration is increased to above ~ 6 µg/10 ng DNA (Figs. 5a and 5b). Binding to the G-rich site is insensitive to the presence of 0.5 µg of competitor DNA, whereas the 3' binding site (site B+B') is no longer protected from DNAse I cleavage (Fig.4), indicating that at least three adjacent sequences are bound by the protein fraction and the binding activities can be distinguished by their concentrations and/or affinity for specific DNA sequence (see Figs.4 and 5). Additional weak binding sites can be occasionally resolved at even higher protein concentrations (25 µg/20 ng DNA) and, for example as shown in Fig.5A (site C), one site maps to position -92 to -66, which includes the CAAT concensus sequence (33).

The tissue specificity of the binding activity(ies) was examined using a 0.3 C4 fraction of nuclei from 3 week old chicken thymus. Although the total amount of thymus C4 protein recovered was twice that from an equivalent number of mature erythrocyte nuclei, filter binding experiments using nuclear-equivalents of thymus C4 protein (0.8 µg protein/150 ng DNA) with end-labelled

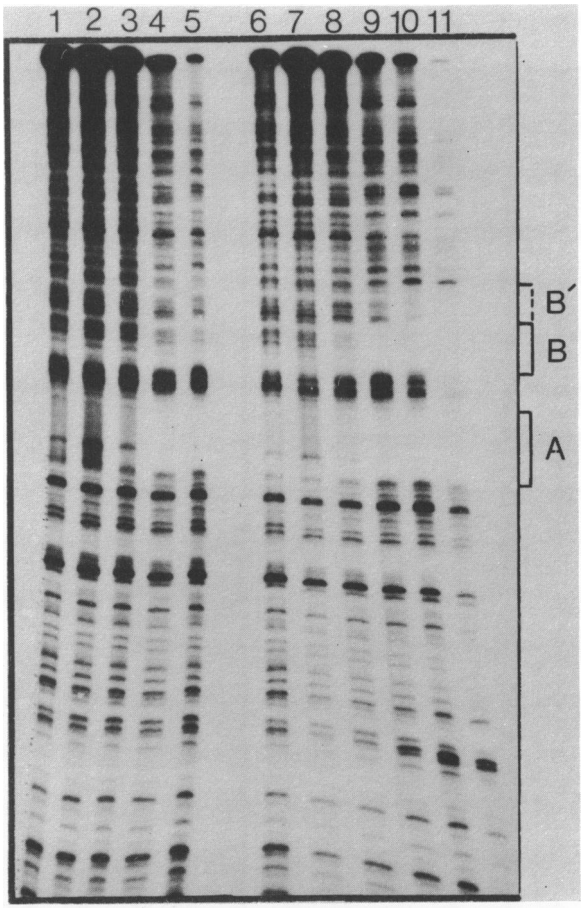


Figure 4:
Footprint analysis of the erythrocyte C4 protein fraction bound to the 5' flanking sequence of the chicken β^A-globin gene.
The coding strand of the $pB^A650.1$ insert (Fig.1) was 5' end-labelled at the HindIII site. The fragment (~10ng) was incubated with increasing mature erythrocyte C4 protein (prepared from pooled 0.14M and 0.3M NaCl nuclear extracts) in the presence (lanes 1-5) or absence (lanes 6-11) of 500ng of competitor E.coli DNA in a final volume of 30µl in RSB (10 mM Tris-Cl (pH7.4), 10 mM NaCl, 3 mM $MgCl_2$) lanes 1 and 6, or with binding buffer (lanes 2-5 and 7-11). Lanes 1,2,6 and 7: no protein controls; lanes 3 and 8: 125ng of C4 protein; lanes 4 nd 9: 1µg of C4 protein; lanes 5 and 10: 2.5µg of C4 protein; lane 11: 5µg of C4 protein.

chicken β^A-globin gene fragments did not exhibit any sequence specificity (data not shown). Using excess thymus C4 protein (~ 4 µg protein/20 ng DNA), footprint analysis (shown in Fig.5) did reveal that the two main erythrocyte

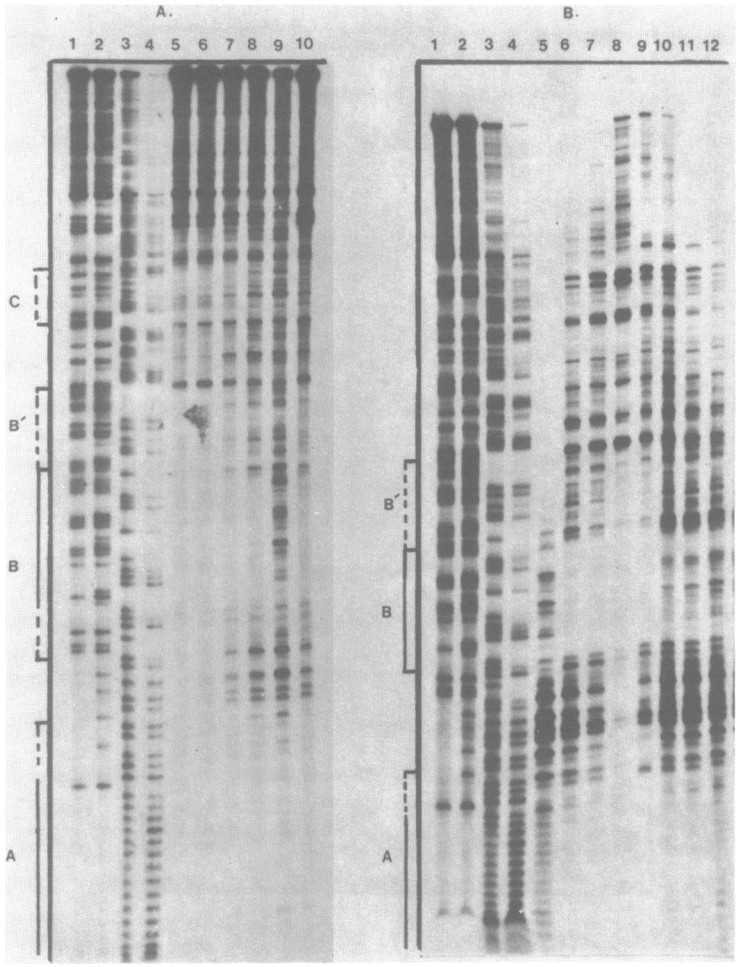
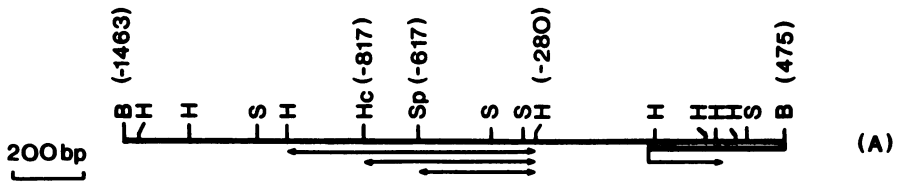


Figure 5:
Footprint analysis of the thymus and eruthrocyte C4 protein fractions bound to the 5' flanking sequence of the chicken β^A-globin gene.
The coding strand of $pB^A650.1$ (Fig.1) was 5'-end labelled at the PstI site, and the PstI-EcoRI fragment gel purified after secondary EcoRI digestion. The fragment was either sequenced by the Maxam and Gilbert procedure, or ~20ng incubated with increasing thymus or erythrocyte C4 protein in a final volume of 100 µl in storage buffer.
(A) Lanes 1-4: C, C+T, G+A and G sequencing reactions respectively. Lane 9: no protein control; lanes 5-8: 25µg, 12.5µg, 5 and 0.5µg of mature erythrocyte C4 protein respectively; lane 10: 4µg of thymus C4 protein (prepared from one 0.3M NaCl nuclear extract).
(B) Lanes 1-5: C, C+T, G+A and G sequencing reactions respectively. Lane 5: no protein control; lanes 6-8: 0.5µg, 1.25µg and 10µg of mature erythrocyte C4 protein respectively; lanes 9-12: 4µg, 2µg, 1µg and 0.4µg of thymus C4 protein respectively.



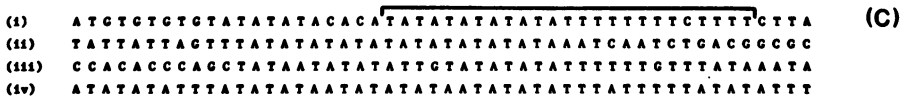


Figure 6A:
Restriction map of the 5' flanking and coding region of the human β-globin gene (pBF5.6 (27,36)). Numbers next to the restriction sites represent their position (in bp) relative to the mRNA CAP site and the arrow shows the direction of transcription. B: BamH1, H: Hinf1, S: SpH1, Hc: Hincll.
Figure 6B:
Sequence of the pBF5.6 Hinf1-HinF fragment in the 5' flanking region of the human β-globin gene (position -817 to -280). The 51 bp alternating purine/pyrimidine sequence is underlined. The sequence protected from DNAse I cleavage by binding to the erythrocyte and thymus C4 protein fractions in footprint analyses (Fig.8) is boxed.
Figure 6C:
Comparison of the human β-globin gene $(TA)_6(T)_8C(T_4)$ sequence and its 5' flanking sequence to (i) similar sequences; (ii) the first intron of the chicken c-myc gene (35); (iii) and (iv) two sequences present in the non-coding strand of the human N-ras gene (30).

C4 protein binding sites (sites A and B+B') were protected from DNAse I cleavage to approximately the same extent as by 500 ng of mature erythrocyte C4 protein under comparable conditions (see Fig.3 and 5). This provides indirect evidence that the binding activity(ies) are in the order of 4 fold more abundant in mature erythrocyte nuclei, and by extrapolation at least 10 fold more abundant in embryonic erythrocyte nuclei, than in thymus nuclei. It must be emphasised that these semi-quantitative estimates assume a direct comparison between filter-binding and footprint analyses.
Filter binding analysis of the erythrocyte C4 protein fraction with other

cloned chicken β-globin sequences (pB2H9, pB2EH325 and pB2EH2, see Fig.1) yielded negative results (data not shown) indicating that the β^A-globin gene binding activity is specific for one site in the 10 kb of the β-globin domain analysed. The erythrocyte protein fractions eluted from the DNA-cellulose at lower (100 mM, C3 fraction) or higher (500 mM, C5 fraction) ammonium sulphate concentrations did not exhibit specific binding to any of the β-globin domain probes available (see Fig.1). Additional filter binding analyses of C3, C4 and C5 protein fractions with a cloned genomic chicken histone gene cluster (pCH3.3E, contains H1, H2A and H2B genes (29)) and with a cloned genomic chicken c-myc gene (pC-c-myc e (28)) were also negative, indicating that the filter binding assay detected specific binding to ∿ 80 bp out of approximately 25 Kb of cloned chicken genomic DNA analysed. Sequence analysis of the chicken c-myc gene (35) shows that there are two polypurine sequences in the 5' flanking sequence and in the first intron (see Fig.3B), which closely resemble the $(G)_{18}CGGGTGG$ sequence bound in the chicken β^A-globin gene. This suggests that specific binding to the 5' flanking region of the β^A-globin gene is not exclusively dependent on the poly(dG) homopolymer but depends on flanking sequences and/or the presence of the two other binding sites(s).

Sequence-specific binding to human gene sequences

The possibility of additional activities was investigated by filter binding analysis of the erythrocyte and thymus C4 protein fractions with restriction fragments of three cloned human genomic sequences: (i) the human β-globin gene (pBF5.6, see Fig.6 and (22)), (ii) the human c-myc gene (pH-c-myc (34)) and (iii) the human N-ras gene (pAT7.0 (30)).

As shown in Fig.7A and Table 1, both the thymus and erythocyte C4 fractions bind a 1.0 Kb Pstl fragment in the N-ras gene. Two human c-myc gene sequences (see Fig.7B) are retained by both protein fractions: an 840 bp Sstl-Pvull fragment in the first intron, and a 1400 bp Pvull-EcoRI fragment spanning the second exon (34). Neither of the known sequences in the N-ras or c-myc fragments bears any resemblance to the filter bound sequence(s) in the chicken β^A-globin gene. There is a polypurine sequence in the human c-myc 5' flanking regions (Fig.3B (iv)) but, as with chicken c-myc (34), this is unretained in a filter binding assay. The N-ras and c-myc sequences are bound by both the thymus and erythocyte C4 fractions (Table 1) whereas only the erythrocyte C4 fraction binds the chicken β^A-globin gene, suggesting at least two unrelated binding activity(ies).

On the other hand, specific binding to a 437 bp Hinfl-Hincll fragment (see

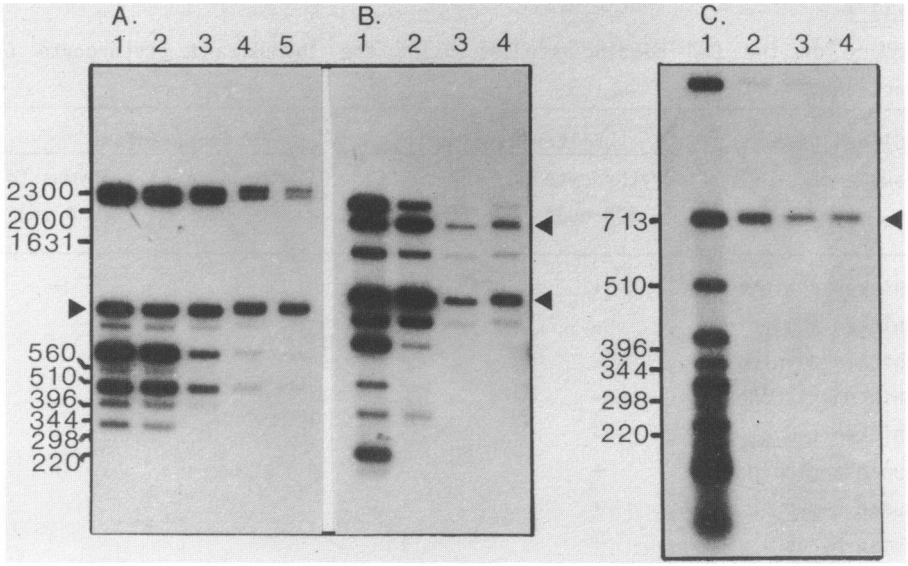


Figure 7:
Filter binding analysis of the erythrocyte C4 protein fraction bound to human gene sequences.
(A) Filter binding to an end-labelled EcoRI, BamHI and PstI triple digest of the human N-ras clone pAT7.0 (30). Lane 1: control DNA; lanes 2-5: 1µg of mature erythrocyte C4 protein was incubated with 150 ng of restriction fragments in the presence of 0,2,4 and 10µg of competitor E.coli DNA respectively.
(B) Filter binding to an end-labelled EcoRI, PstI and PvuII triple digest of the human c-myc clone pH c-myc (34). Lane 1: control DNA; lanes 2-4: 1µg of mature erythrocyte C4 protein was incubated with 150ng of end-labelled restriction fragments in the presence of 0,5 and 10µg of competitor E.coli DNA respectively.
(C) Filter binding to an end-labelled HinfI and BamHI double digest of the human β-globin clone pBF5.6 (see Fig.6). Lane 1-4: 1µg of erythrocyte C4 protein and 150ng of end-labelled restiction fragments were incubated in the presence of 0,2,4 and 10µg of competitor E.coli DNA respectively.

Fig.6) in the 5' flanking region of the human β-globin gene (position -817 to -230) was only detected in mature and embryonic C4 protein fractions (see Fig. 7C and Table 1). Footprint analysis of the mature erythrocyte C4 protein bound to an SphI-HinfI fragment (^{32}P)-ATP end-labelled at the SphI site (Fig. 8) shows that a $(TA)_6(T)_8C(T)_4$ sequence (position -516 to -540 (36) and Fig. 6B) is the major site selectively protected from DNAse I cleavage. This site occurs at the 3' end of a 48 bp alternating purine/pyrimidine sequence (see Fig.6B), and as only 25% of that sequence is bound, it suggests that the 3' $(T)_8C(T)_4$ sequence is required for binding. In agreement with this, a $(TA)_{11}$ sequence present in the first intron of the chicken c-myc clone (pC-c-myc,

Table 1

Summary of the DNA binding studies using the thymus and erythrocyte C4 fractions.

Cloned genomic sequence	Filter binding		Footprinting	
	Erythrocyte C4 (∿ 500 ng)*	Thymus C4 (∿ 800 ng)*	Erythrocyte C4 (∿ 500 ng)*	Thymus C4 (∿ 4 µg)*
chicken β^Aglobin	+	-	+	+
chicken β^Hglobin	-			
chicken β^{∞} globin	-	-		
chicken histone	-	-		
chicken c-myc	-	-		
human β-globin	+	-	+	+
human c-myc	+	+		
human N-ras	+	+		

* Figures represent an estimate of the protein (ng) used to bind 150 ng of DNA in filter binding studies or 10 ng of DNA in footprinting studies. Sequence specific binding is denoted by (+).

Fig.6C (ii) and (35)) is not filter bound under comparable conditions. The human N-ras 1 Kb Pst1 fragment contains sequences comparable to that in the human β-globin gene (Fig.6C), but the observation that the thymus C4 protein fraction binds to N-ras but not to human β-globin sequences, under the same conditions, indicates that these are different binding activities.

As with the chick β^A-globin, the thymus C4 protein (0.8 µg protein/150 ng DNA) does not appear to bind to the human β-globin gene (Table 1). However, footpring analysis using excess thymus protein (4 µg/20 ng DNA) revealed the selective protection of the $(TA)_6(T)_8C(T)_4$ sequence from DNAase I cleavage (Fig.8). Thus the binding of the C4 thymus and erythocyte protein fractions to the chicken β^A- and human β-globin 5' flanking sequences, as defined by filter binding and footprinting analyses, occur under similar conditions (Figs. 2,4,5,7 and 8).

DISCUSSION

The results presented above suggest that proteins eluted from chicken thymus and erythrocyte nuclei with 0.15M and/or 0.3M NaCl have several classes of sequence specific DNA-binding activities, some of which have been characterised by filter binding and footprint analyses with cloned homologous and

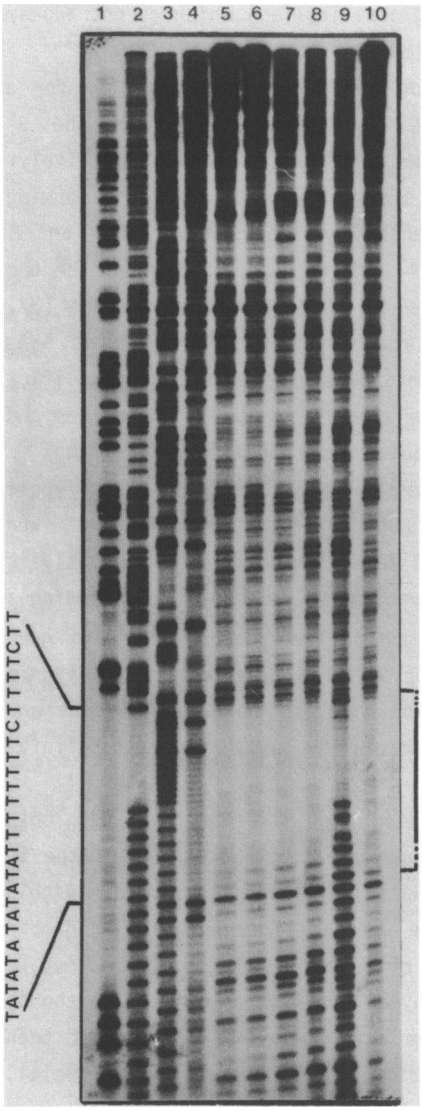


Figure 8:
Footprint analysis of thymus and erythrocyte C4 protein bound to the 5' flanking sequence of the human β-globin gene.
The coding strand of the pBF5.6 insert (Fig.6) was 5' end-labelled at the SpH1 site and the SpH1-BamHI fragment gel purified after secondary BamHI digestion. Lanes 1-4: G, G+A, T+C and C sequencing reactions; lane 9: no protein control; lanes 5-8: 25, 12.5, 1.25 and 0.5μg of erythrocyte C4 protein incubated with 20 ng of fragment; lane 10: 4μg of thymus C4 proteins incubated with 20ng of fragment. The site protected from DNAse I digestion is shown in Fig.6.

heterologous genomic DNA sequences.

One class of sequence-specific DNA-binding activity(ies), in both thymus and erythrocyte nuclei, is detectable by filter binding analyses to sequences present in genomic human N-*ras* (30) and c-*myc* (34) genes. The human N-*ras* sequence recognised is extremely A+T rich (A. Hall, personal communication), but is not obviously related to the two sequences bound in the c-*myc* clone.

The functional significance of these binding activity(ies) which do not appear to be tissue specific remains to be elucidated.

A second class of binding activity(ies) exhibits sequence specificity for a $(G)_{18}C(G)_3T(G)_2$ sequence or an $(AT)_6(T)_8C(T)_4$ sequence found in the 5' flanking regions of the chicken β^A- and human β-globin genes respectively. As both these sequences occur at, or near, sites which are capable of forming altered secondary DNA structures as evidenced by their sensitivity to S1 nuclease in supercoiled plasmid DNA (9,22-24), it is possible that either one DNA-binding activity is recognising a particular secondary DNA structure (37, 38) or two activities specifically bind distinct primary DNA sequences. As other polypurine or alternating purine/pyrimidine sequences are not specifically retained in filter binding analyses, we conclude that 3'- boundaries and sequences flanking the binding sites are required for specific recognition.

This class of binding activity(ies) is most abundant in 15 day embryonic erythrocyte nuclei, is ∿ 5-fold less abundant in adult erythrocytes, and comparatively depleted in chick thymus (but detectable by footprint analysis using a much higher protein : DNA ratio (w/w) than used in filter binding). As Emerson and Felsenfeld (19) could not detect the chicken β^A-globin gene sequence binding activity in chick brain or 9 day old (primitive) embryonic erythrocyte nuclei, it suggests that the binding activity is both tissue and developmentally regulated, being most abundant in the early stages of definitive erythrocyte maturation.

The significance of the proteins binding to the chicken β^A-globin $(G)_{18}C(G)_3T(G)_2$ sequence is not clear although the sequence occurs at the 5' boundary of the developmentally regulated DNAse I hypersensitive domain within the nucleosome free chromatin region of the definitive erythrocyte β^A-globin gene (8). It has been suggested that this class of binding protein(s) may be involved in the generation of this DNAse I hypersensitive site (19). Although this activity was not detected in chicken brain nuclei (19) it has been possible to detect the specific binding activity in chicken thymus nuclei, albeit at least 4 fold lower levels. This, together with an apparent preference for specific altered secondary DNA structures, suggests that the protein factor(s) may not be directly responsible for the generation of DNAse I sensitive domains but may bind to, and possibly maintain, pre-existing altered free DNA or chromosomal DNA secondary structures. However, it was not possible to detect sequence specific binding to other sequences on chicken β-globin domain, namely the β^ρ and β^H genes and flanking sequences, despite the fact that they contain sequences which are hypersensitive to S1

nuclease in supercoiled plasmids and domains which are developmentally regulated and have endonuclease hypersensitive sites in erythrocyte chromatin (6-9). No sequence specificity was exhibited for a cloned genomic chicken histone gene cluster (H1, H2A and H2B genes (29)), nor for a cloned genomic chicken c-myc gene (28), suggesting that the DNA binding activity is not generally associated with DNAse I hypersensitive sites in chromatin, whether developmentally regulated or otherwise. The $(TA)_6(T)_8C(T)_4$ sequence in the 5' flanking region of the human β-globin gene has not been identified as a major DNAse I hypersensitive site in human erythroid chromatin (41), but the functional significance of the sequence in vitro is suggested by the identification of a β-globin gene in a β^+ thalassemia patient in which an ATA trinucleotide had been substituted for a T residue at nucleotide-530 bp upstream from the mRNA CAP site (42). This substitution, in an otherwise apparently functional β-globin gene sequence, occurs in the middle of the chicken erythocyte C4 protein binding site (see Fig.6). Similarly, sequences between 344 bp and 1314 bp upstream from the mouse β-major globin gene have been shown to inhibit transcription of the gene in transfected mouse fibroblasts (43). This region contains a set of imperfect tandem repeats of $(ACAT)_N$ tetranucleotides and could possibly form Z-DNA. Thus, these types of sequences could be cis negative regulatory elements which might be derepressed by specific DNA-binding protein(s).

A third class of sequence-specific DNA binding activities was detected by footprint analysis of the 5' flanking region of the β^A-globin gene, and are present in lower amounts or have a lower affinity for the specific sequence(s) than the other binding activities. At least two binding sites were detected between the $(G)_{18}C(G)_3T(G)_2$ sequence and the ATAA consensus sequence (Fig.3). The major binding site, detected by binding to both thymus and erythrocyte nuclear protein fractions, commences 15 bp downstream from the G-rich binding site and extends to nucleotide -137. With increasing protein in the absence of competitor DNA, this footprint expands to nucleotide -115 (see Fig.3). The CACCC consensus sequence, commonly found in the 5' flanking region of β-globin genes (44) occurs at the boundary between the two binding sites (Fig.3). Downstream of this binding site is a GCCCGGGA...AGGGCCCG inverted repeat which is also hypersensitive to S1 nuclease in supercoiled plasmid DNA (21). These binding sites (nucleotide -115 to -161) occur within the developmentally regulated DNAse I hypersensitive domain in definitive erythrocyte chromatin (8), and with the G-rich binding site (between nucleotides -177 and -201, see Fig.3) occupy approximately 70% of that domain whch is endonuclease hypersens-

itive and nucleosome free in chicken erythrocyte β^A-globin gene chromatin (8). This domain contains consensus sequences which have been shown to be important for the in vivo expression of other transfected globin genes, namely the CAAT box and the CACCC consensus sequence which have been found to be important for the transcription of rabbit β-globin genes, and is present in many other globin genes (for review see (44)). These binding activities could therefore be similar to transcription factors like the SP1 protein which binds to the CCGCCC repeats upstream from the SV40 early promoter (16) and the drosophila proteins which bind to the heat-shock element of the drosophila heat shock genes (11-13).

We are currently attempting further purification of the protein factors which together with in vitro transcription studies may lead to a better understanding of their function in vivo.

ACKNOWLEDGEMENTS

We thank Drs. A. Hall, J. Welles, K. Bister, A. Malcolm and D. Engels for providing genomic clones, and Drs. K. Dudley, I. McKay and P. Little for help and advice during these studies. These studies were supported by grants from the Medical Research Council and the Cancer Research Campaign.

REFERENCES

1. Villeponteau,B. and Martinson,H.G. (1981) Nucleic Acids Res. 9, 3731-3746.
2. Dolan,M., Sugarman,B.J., Dodgson, J.B. and Engel,J.D. (1981) Cell 24, 669-677.
3. Villeponteau,B., Landes,G.M., Pankratz,M.J. and Martinson,H.G. (1982) J.Biol.Chem. 257, 11015-11023.
4. Landes,G.M., Villeponteau,B., Pribyl,T.M. and Martinson,H.G. (1982) J.Biol.Chem. 257, 11008-11014.
5. Weintraub,H. and Groudine,M. (1976) Science 93, 848-853.
6. Stadler,J., Larsen,A., Engel,J.D., Dolan,M., Groudine,M. and Weintraub,H. (1980) Cell 20, 451-460.
7. Groudine,M. and Weintraub,H. (1981) Cell 24, 393-401.
8. McGhee,J.D., Wood,W.I., Dolan,M., Engel,J.D. and Felsenfeld,G. (1981) Cell 27, 45-55.
9. Larsen,A. and Weintraub,H. (1982) Cell 29, 609-622.
10. Parker,C.S. and Topol,J. (1984) Cell 36, 357-369.
11. Wu,C. (1984) Nature 309, 229-234.
12. Jack,R.S., Gehring, W.J. and Brack,C. (1981) Cell 24, 321-331.
13. Parker,C.S. and Topol,J. (1984) Cell 37, 273-283.
14. Grover,R., Kennedy,N., Skroch,P., Hynes,N.E. and Ponta,H. (1984) Biochim.Biophys. Acta 781, 1-6.
15. Dilworth,S.M., Cowie,A., Kamen,R.I. and Griffin,B.E. (1984) Proc.Natl.Acad.Sci.USA 81, 1941-1945.
16. Dynan,W.S. and Tjian,R. (1983) Cell 35, 79-87.
17. Shastry,B.S., Honda,B.M. and Roeder,R.G. (1984) J.Biol.Chem. 259, 11373-11382.

18. Siebenlist,U., Hennighausen,L., Battey,J. and Leder,P. (1984) Cell 37, 381-391.
19. Emerson,B.M. and Felsenfeld,G. (1984) Proc.Natl.Acad.Sci.USA 81, 95-99.
20. Mulvihill,E.R., Le Pennee,J.P. and Chambon,P. (1982) Cell 28, 621-632.
21. Evans,T., Schon, E., Gora-Maslak, G., Patterson,J. and Efstratiadis,A. (1984) Nucleic Acids Res. 12, 8043-8058.
22. Nickol,J.M. and Felsenfeld,G. (1983) Cell 35, 467-477.
23. Weintraub,H. (1983) Cell 32, 1191-1203.
24. Cockerill,P.N. and Goodwin, G.H. (1983) Biochem.Biophys.Res.Commun. 112, 547-554.
25. Vieira,J. and Messing,J. (1982) Gene 10, 259-268.
26. Dodgson,J.B., Strommer,J. and Engel,J.D. (1979) Cell 17, 879-887.
27. Jackson,I.J., Freund,R.M., Wasylyk,B., Malcolm,A.D.B. and Williamson,R. (1981) Eur.J.Biochem. 121, 27-31.
28. Robins,T., Bister K., Garon,C., Papas,T. and Duesberg,P. (1982) J.Virol. 41, 635-642.
29. D'Andrea,R., Harvey,R. and Wells,J.R.E. (1981) Nucleic Acids Res. 9, 3119-3128.
30. Brown,R., Marshall,C.J., Pennie,S.G. and Hall,A. (1984) EMBO J. 3, 1321-1326.
31. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) in Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory.
32. Maxam,A.M. and Gilbert,W. (1977) Proc.Natl.Acad.Sci.USA 74, 560-564.
33. Dolan,M., Dodgson,J.B. and Engel,J.D. (1983) J.Biol.Chem. 258, 3983-3990.
34. Colby,W.W., Chen,E.Y., Smith,D.H. and Levinson,A.D. (1983) Nature 301, 722-725.
35. Shih,C.-K., Linial,M., Goodenow,M.M. and Hayward,W.S. (1984) Proc.Natl. Acad.Sci.USA 81, 4697-4701.
36. Moschonas,N., de Boer,E. and Flavell,R.A. (1982) Nucleic Acids Res. 10, 2109-2120.
37. Arnott,S., Chandrasekaran,R., Birdsall,D.L., Leslie,A.G.W. and Ratliff, R.L. (1980) Nature 283, 743-745.
38. Cantor,C.R. and Efstratiadis,A. (1984) Nucleic Acids Res. 12, 8059-8072.
39. Compton,J.G., Schrader,W.T. and O'Malley,B.W. (1983) Proc.Natl.Acad.Sci. USA 80, 16-20.
40. Borgmeyer,U., Nowock,J. and Sippel,A.E. (1984) Nucleic Acids Res. 12, 4295-4311.
41. Groudine,M., Kokwi-Shigematsu,T., Gelinas,R., Stamatoyannopoulos,G. and Papayannopoulou,T. (1983) Proc.Natl.Acad.Sci.USA 80, 7551-7555.
42. Semenza,G.L., Delgrosso,K., Poncz,M., Malladi,P., Schwartz,E. and Surrey, S. (1984) Cell 39, 123-128.
43. Gilmour,R.S., Spandidos,D.A., Vass,J.K., Gow,J.W. and Paul,J. (1984) EMBO J. 3, 1263-1272.
44. Collins,F.S. and Weissman,S.M. (1984) in Progress in Nucleic Acid Research and Molecular Biology, Vol 31, pp. 315-462, Academic Press.