Volume 15 Number 19 1987 Nucleic Acids Research

Multiple transcription factors interact with the adenovirus-2 EII-late promoter: evidence for a novel CCAAT recognition factor

C.R.Goding*, S.M.Temperley and F.Fisher

Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TL, UK

Received August 5, 1987; Accepted September 7, 1987

ABSTRACT

Multiple cellular transcription factors have been shown to interact with the upstream region of the adenovirus-2 EIIa-late promoter. One of these factors recognises each of the three CCAAT motifs present in the EIIL promoter at positions -72, -135 and -229, as well as the CCAAT elements in the rat albumin and herpes virus thymidine kinase promoters. A mutation known to reduce thymidine kinase promoter activity *in vivo* and *in vitro* abolishes binding of the factor, termed CCAAT recognition factor (CRF), which appears to be distinct from previously identified CCAAT factors. In addition, another protein, termed upstream factor II (USFII), shares binding sites at position -110 in the EIIL promoter and in the c-*fos* enhancer adjacent to the serum regulatable element. The recognition site for USFII is also found in the c-*fos* promoter and in the adenovirus early region EIV and EIIa-early promoters. An Sp1 recognition site has also been identified at position -41, and the binding sites for Sp1, USFII and CRF are all required for efficient EIIa-late promoter function. Finally, an additional factor recognising the consensus GGGGGGNT has been detected.

INTRODUCTION

Control of transcription plays a key role in regulation of gene expression during development and the cell cycle and in the cellular response to hormones, growth factors, heat-shock and other external stimuli. For many genes the sequence requirements for constitutive and regulated initiation of transcription have been defined, with the transcriptional status of a given promoter being conferred by the combination and spatial arrangements of its cis-acting elements. Development of techniques designed to examine the interactions of these cis-acting elements with sequence-specific DNA-binding proteins has led to the identification and purification of factors such as Sp1 (1) and CTF (2) required for efficient constitutive transcription from a number of eukaryotic viral and cellular promoters. It is clear however that different factors may share common promoter elements; the CCAAT motif of the herpes simplex virus thymidine kinase promoter (HSV TK), for example, is recognised both by CTF (2) and by CBP (3). Thus it is possible that some degree of transcriptional regulation could be mediated through such

© IRL Press Limited, Oxford, England.

constitutive promoter elements. Analysis of viral promoters differentially expressed during infection provides an opportunity to examine the cis- and trans-acting requirements for both constitutive and regulated gene expression. In this respect the adenovirus EIIa transcription unit, encoding a 72,000 molecular weight protein primarily involved in viral DNA replication, provides an excellent system for studying regulated eukaryotic transcription. Early in infection, before the onset of DNA replication at about 6 hours post-infection, transcription is under the control of the EIIa-early (EIIaE) promoter located at about 75 map units. After DNA replication has begun transcription from the EIIaE promoter is repressed and a second promoter EIIa-late (EIIL) located at 72 map units is activated (4). In contrast to the EIIaE promoter, EIIL is not induced by products of the adenovirus immediate early EIa gene (5), and one report suggests that the product of the EIa 12S mRNA may repress EIIL expression (6).

To gain an understanding of the differential regulation of this transcription unit we have undertaken an analysis of the cellular factors recognising the EIIL promoter. Using a HeLa cell nuclear extract and a combination of DNAse 1 footprinting (7), a sensitive gel electrophoresis-DNA binding assay (8, 9), competition and methylation interference experiments, we have identified multiple cellular factors binding independently to the EIIL promoter. One of these factors recognises the CCAAT motifs present in the EIIL, rat albumin and HSV TK promoters and may represent a novel CCAAT binding activity. The results are discussed in the light of recent evidence as to the cis-acting requirements for EIIL transcription.

MATERIALS AND METHODS

Recombinant Plasmids and Probes

Plasmid pLi is essentially the same as pL (5) but with the EIIL sequences between the Fnu DII sites at +32 and -264 in the inverse orientation with respect to the rabbit β-globin coding sequences (-9 to +1650). Probe A contains EIIL sequences from the Rsa I site at position -32 to the Hae III site at position -134 cloned into the Sma I site of the pUC19 polylinker. In addition probe A has the small Hae III fragment (-122 to -134) present between the Sma I and Rsa I sites. Probe C contains EIIL sequences between the Fnu DII site at position -264 and the Sac I site at position -220 and was constructed by inserting the Sac I fragment from pLi into the Sac I site of the pUC19 polylinker. Probe B contains EIIL sequences between the Hae III site at position -134 and the Sac I site at position -220 inserted between the Sma I and Sac I sites of the pUC19 polylinker. Probe V contains a synthetic oligonucleotide corresponding to EIIL sequences between positions -186 and

-213 cloned into the BamHI site of the pUC19 polylinker.

All oligonucleotide probes were synthesised on an Applied Biosystems model 381A synthesiser and were hybridised under standard conditions prior to use.

Gel electrophoresis DNA-binding and methylation interference assays

Nuclear extracts were prepared exactly as described (10) and protein concentrations were determined by the method of Bradford (11).

Binding reactions were performed in a 20μl volume containing 0.1 to 6μl of nuclear extract 25mM hepes pH7.9, 1mM EDTA, 5mM DTT, 150, KCl, 10% Glycerol (10,000 c.p.m, -0.5ng) labelled DNA fragment and 1-2μg poly(dI-dC).poly(dI-dC) and incubated at room temperature for 20′ before loading directly onto an 8% acrylamide gel (44:1 acrylamide:bis) and electrophoresis in 0.5 X TBE at 10 v/cm. Fragments for competition studies were prepared by sucrose gradient fractionation and ethanol precipitation. Labelled fragments for the DNA-binding assay or footprinting studies were prepared by 3′-end-labelling with Klenow and isolated by electrophoresis on a 4% acrylamide gel followed by electroelution and ethanol precipitation. For methylation interference experiments DNA was partially methylated *in vitro* with DNA as described (12) before being used in the gel electrophoresis DNA-binding assays described above. Retained and unbound DNA was electroeluted, ethanol precipitated and cleaved with piperidine before being subjected to electrophoresis and autoradiography.

DNAse 1 protection assays

Footprinting reactions were carried out in 50μl in a buffer containing 25mM Hepes pH 7.9, 1mM $MgCl_2$, 5mM DTT, 5mM NH_4SO_4,60mM KCl, 12% glycerol, 2μg poly(dI-dC).poly(dI-DC) AND 10,000 c.p.m labelled DNA fragment. After incubation with extract at room temperature for 30′, 5μl of an appropriate concentration of DNAse 1 was added for 60 seconds before the reactions were stopped by addition of 30μl 3M NH_4Ac, 20mM EDTA. The samples were then extracted once each with phenol and chloroform before being ethanol precipitated, and resuspended in 80% formamide prior to electrophoresis on 8% acrylamide, 50% urea sequencing gels.

RESULTS

Detection and footprinting of factors that bind to the EIIL promoter

In the absence of any previous data on cis-acting requirements for EIIL transcription we examined directly the interactions between cellular factors and the EIIL upstream sequences. Identification of factor binding sites would enable subsequent targeted mutagenesis of the promoter in the viral genome and purification of the factors should eventually enable the promoter to be analysed in a reconstituted *in vitro* transcription system. As a source

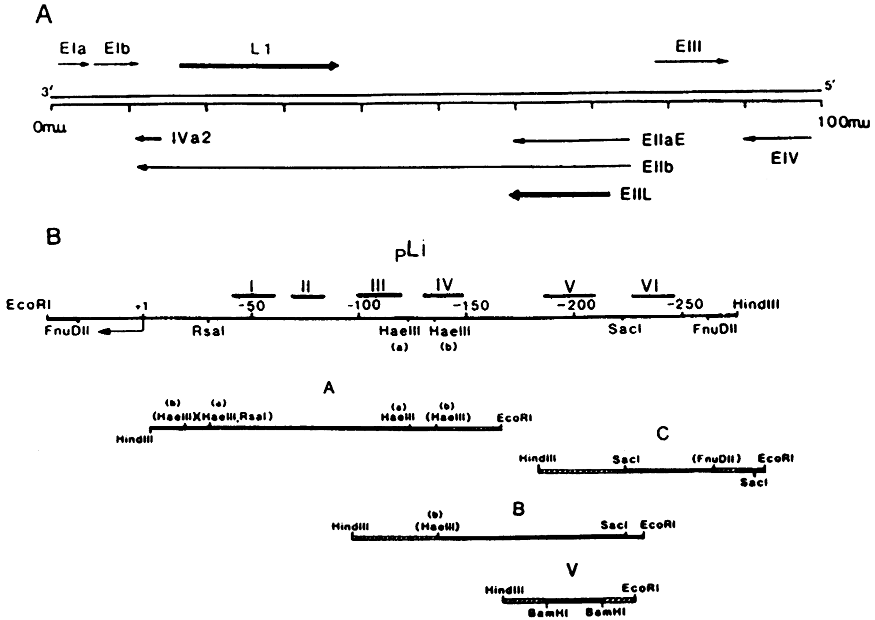

Figure 1A. Simplified map of the adenovirus genome showing relative positions of primary transcripts. Transcripts from the major late promoter (L1) and EIIL promoter are represented as thick lines.
Figure 1B. Map of recombinants. The Fnu DII fragment from position +32 to -264 containing the EIIL promoter is shown with the direction of transcription indicated by the arrow from the cap-site at +1. The whole of this Fnu DII fragment is present in clone pLi. The EcoRI-HindIII fragments from EIIL subclones A, B, C and V used in the gel electrophoresis DNA-binding assays and methylation interference experiments are shown below. The filled bars represent EIIL sequences derived from pLi and the hatched lines pUC19 polylinker flanking sequences. Restriction sites in parentheses were lost during cloning.

of cellular factors we used an HeLa cell nuclear extract prepared as described (10). To determine whether the extracts contained factors that recognised the EIIL upstream sequences and to follow the purification of any such factors, we used the sensitive gel electrophoresis DNA-binding assay (8, 9). Fragments of the EIIL promoter were subcloned into the pUC18 polylinker (Figure 1b, probes A-C) and 3′ end-labelled at the Hind III site and the EcoRI-Hind III fragments isolated. Complexes were formed with each of the probes (not shown) and the different migration patterns of the complexes suggested that multiple different factors might be recognising the EIIL sequences.

To localise the binding sites of the factors detected by the gel electrophoresis DNA-binding assay and to determine whether these binding sites

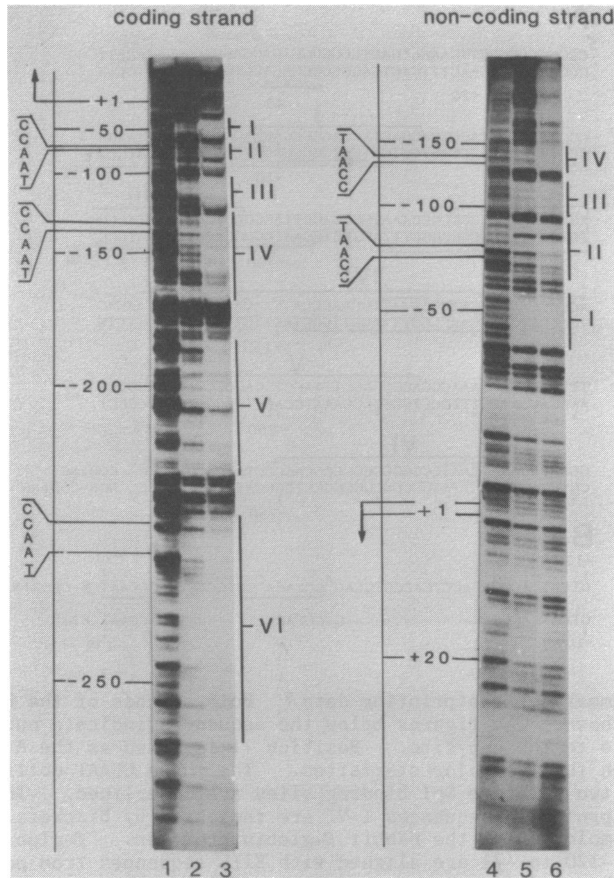

Figure 2. DNAse 1 footprinting analysis of EIIL promoter-factor interactions. Lanes 1-3, coding-strand of pLi was labelled at the 3′ end and incubated with: lane 1, no extract; lane 2, 6 µl extract, or lane 3, 12 µl extract, before DNAse 1 digestion and analysis on an 8% sequencing gel. Lanes 4-6 as for lanes 1-3, but with labelled non-coding strand. The positions of the cap-sites are labelled +1 and the arrows indicate the direction of transcription. Also marked are the positions of the CCAAT boxes and the regions of protection, I-VI.

correlated with the putative CCAAT or Sp1 recognition sequences present in the EIIL promoter (see legend to Figure 3), we attempted to analyse the EIIL promoter by DNAse 1 footprinting with crude nuclear extracts.

The EcoRI-Hind III fragment from pLi (Figure 1B) was 3′ end-labelled at the Hind III site and digested briefly with DNAse 1 either in the presence or absence of nuclear extract. Analysis of the products on an acrylamide gel revealed that multiple regions of the EIIL promoter were protected from DNAse

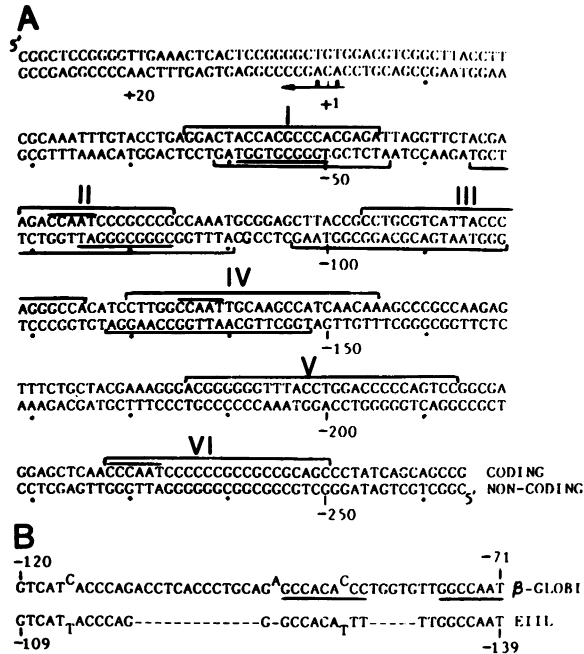

Figure 3. Summary of footprinting data. Both strands of the EIIL FnuDII fragment are shown. The figures below the sequences indicate nucleotide positions relative to the cap-site. Position +1 is taken as the A residue located between the two major startsites. The three CCAAT motifs are overlined and the two putative Spl binding sites are underlined. The approximate extent of the protected sequences I-VI are indicated by brackets.
Figure 3B. Homology with the rabbit β-globin promoter. β-globin sequences from position -120 to -71 are aligned with EIIL sequences from position -109 to -139. Dashes indicate bases present in β-globin but absent in EIIL and mis-matches are indicated by off-set nucleotides. Elements required for efficient β-globin expression are underlined.

1 digestion in the presence of extract (Figure 2, compare lanes 2 and 3 to lane 1). The protected regions, marked I-VI, extend from position-36 to -250 and were reproducibly observed in different experiments using various extract preparations. Digestion of the same fragment labelled on the non-coding strand generated regions of protection (lanes 5 and 6) corresponding to those seen on the coding strand. Comparison of these and other footprinting experiments with sequence ladders of the labelled fragments (not shown) enabled identification of the protected sequences, the approximate localisations of which are shown in Figure 3. The precise location of protection on the non-coding strand of regions V and VI have not yet been determined. No additional protected regions were observed using as a probe a labelled fragment extending between position -413 and position +75.

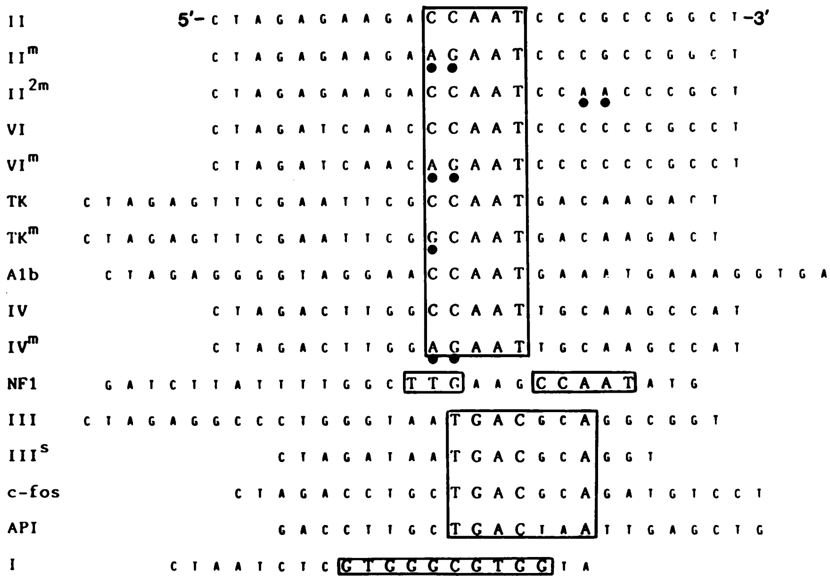

Figure 4. Sequences of oligonucleotides used as probes or competitors. Single-strands are shown 5′ to 3′ with core consensus sequences boxed. Dotted nucleotides indicate positions of point mutations.

Both putative Spl binding sites at positions -41 and -75 (Figure 3) are protected as are two of the three CCAAT motifs, at positions -72 and -135. The third CCAAT box at position -229 is peripheral to region VI, indicating perhaps that another factor recognises these protected sequences. In addition two further protected regions III and V were evident. Although these data provide information as to the location of factor binding sites, they do not enable us to determine whether each of the protected regions detected is due to independent binding of a single factor or whether any one factor binds at multiple sites. We therefore chose to examine in isolation binding sites I and II which each show homology to known transcription factor recognition sequences.

The SV40 21 base pair repeats compete efficiently for binding to region I

Binding site I exhibits a strong homology to other non-consensus Spl recognition sequences (13-15). To examine this region closely a 27 base pair labelled fragment spanning positions -36 to -63 (probe I Figure 4) was used in the gel electrophoresis DNA-binding assay. Nuclear extract was therefore used together with the labelled fragment corresponding to binding site I in a competition assay. In the absence of competitor DNA a single major retained

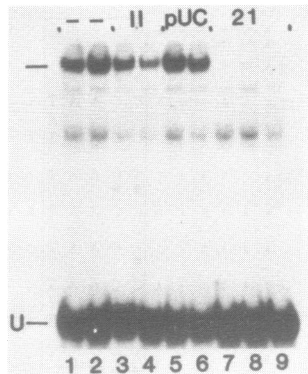

Figure 5. Binding to EIIL region I. A labelled fragment corresponding to probe I (Figure 4) was used in the gel electrophoresis DNA-binding assay, either in the absence of competitor DNA, lanes 1 and 2; or in the presence of 50 or 250ng competitor DNA corresponding to probe II (Figure 4) lanes 3 and 4; pUC18 polylinker, lanes 5 and 6; or 50, 100 or 250ng of a fragment containing the SV40 21 base pair repeats, lanes 7-9 respectively.

band was observed (Figure 5, lanes 1 and 2). In the presence of an excess of the pUC18 polylinker (lanes 5 and 6) no significant competition was observed. In contrast a fragment corresponding to the SV40 21 base pair repeats competed extremely efficiently (lanes 7 - 9), strongly suggesting that the transcription factor Spl recognises the EIIL promoter within binding site I. Only weak competition (2 to 3-fold) was observed using as competitor a subcloned region of the EIIL promoter corresponding to binding site II (lanes 3 and 4) indicating that Spl binds weakly if at all within binding site II despite the apparent homology with other Spl recognition sequences.

In the converse experiment a probe corresponding to binding site II (probe II, Figure 4) formed a complex that was competed for efficiently by the homologous DNA but not by either the pUC polylinker or by the 21 base pair repeats (Figure 6, panel A), supporting the notion that Spl recognises the Spl consensus homology within binding site II weakly if at all. Indeed a double point mutation introduced into this homology (probe II^{2m}, Figure 4) failed to inhibit formation of the complex (Figure 6, panel B), whereas in contrast alterations of the CCAAT motif within binding site II to AGAAT (probe II^{m}, Figure 4) abolished complex formation.

Evidence for a novel CCAAT recognition factor interacting with the EIIL, rat albumin and HSV TK CCAAT elements

It seemed likely that the protection of region II from DNase 1 digestion resulted from the interaction of a cellular factor with the CCAAT motif at position -72. We next asked whether the same factor could also bind to the

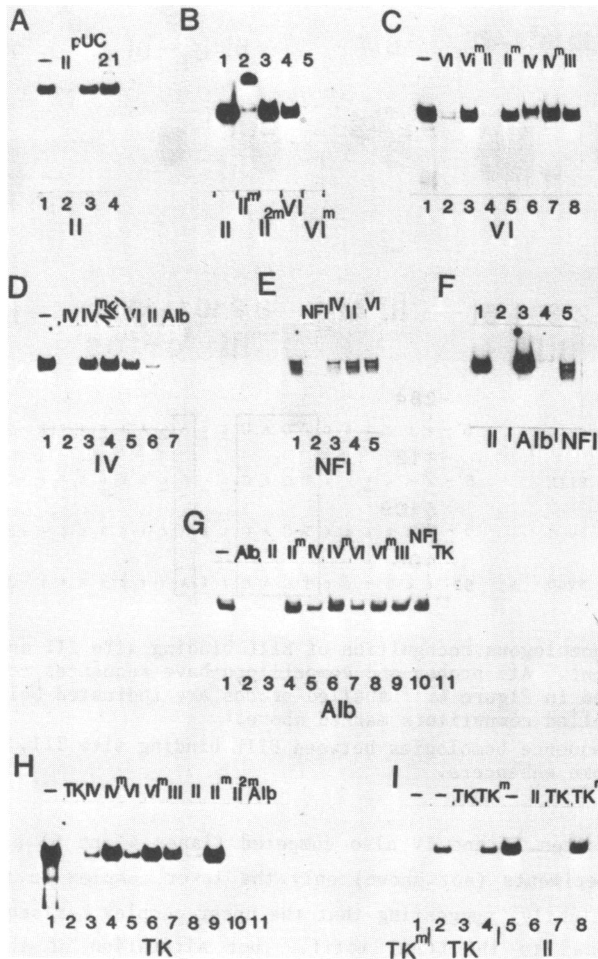

Figure 6. Recognition of CCAAT motifs by a cellular factor. Bands shown correspond to retained DNA-protein complexes, unbound DNA is not shown. All probes are 3′ end-labelled and are indicated below each panel. Unlabelled competitor DNAs are indicated above each panel and are used at approximately 250ng/binding reaction. All probes and competitors correspond to those indicated in Figure 4 or to DNA. In panel F extract was heated to 60°C for 5′ prior to binding in lanes 2 and 4.

CCAAT motifs present within binding sites IV and VI at positions -135 and -229 respectively. Using probe VI we competed for factor binding using double stranded oligonucleotides corresponding to binding sites II, III, IV and VI (see Figure 4 for sequences of probes and competitors). As expected, the homologous probe, VI, competed efficiently, whereas binding site III, lacking a CCAAT box, failed to compete (Figure 6, panel C, lanes 2 and 8 respectively).

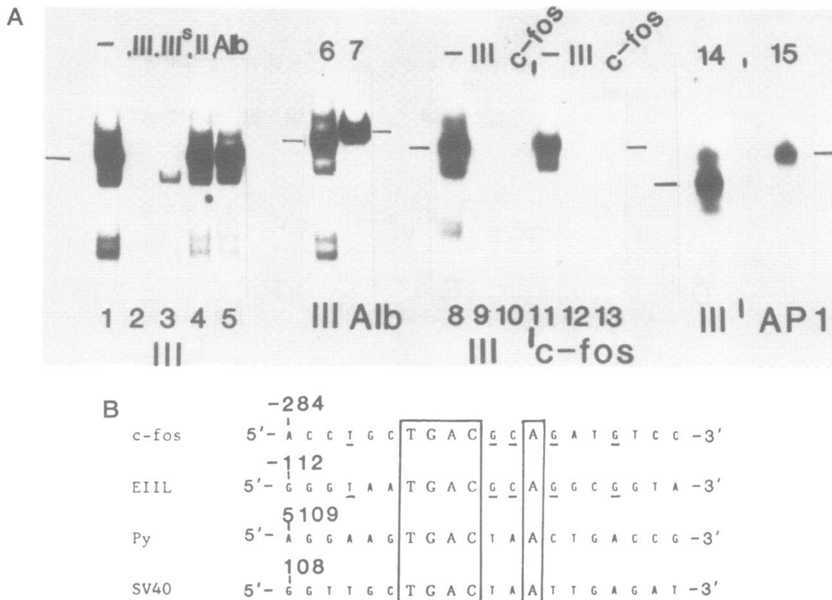

Figure 7A. Homologous recognition of EIIL binding site III and a c-*fos* enhancer element. All probes and competitors have sequences corresponding to those described in Figure 4. Labelled probes are indicated below each panel and the unlabelled competitors marked above.
Figure 7B. Sequence homologies between EIIL binding site III, and the c-*fos*, SV40 and polyoma enhancers.

Both binding sites II and IV also competed (lanes 4 and 6) although in this and other experiments (not shown) only the lower complex in the doublet was competed efficiently, suggesting that the upper complex has sequence requirements additional to the CCAAT motif, but alteration of the CCAAT motifs present within binding sites II, IV and VI to AGAAT (II^m, IV^m and VI^m respectively) abolished competition (lanes 5,7 and 3). Thus under these conditions all three EIIL CCAAT motifs compete for the same factor. The reciprocal experiment, using binding site II as probe, gave identical results (not shown).

In a similar experiment (panel D) using binding site IV as a probe, an identical complex was formed and as expected the homologous sequence competed very efficiently, as did region II, with binding site VI competing around 5-fold less efficiently.

The data presented so far demonstrate that all three CCAAT elements in the EIIL promoter can bind the same factor, which for ease of reference we termed the CCAAT recognition factor (CRF). Although we are using crude nuclear extracts the competition experiments using similar length oligonucleotides give

some indication of the relative affinity of CRF for each of its cognate recognition sites in the EIIL promoter. Bearing in mind that the degree of competition will be dependent on the relative affinities of the probe and the competitor, the data from these experiments and others not shown are consistent with the CCAAT motif within binding site VI having an affinity around 5-fold lower for CRF than regions II and IV.

In a number of other promoters the CCAAT motif is required for efficient initiation of transcription. We wished therefore to determine the promoter specificity of CRF and to establish whether CRF was homologous to either of the two previously identified CCAAT factors, CCAAT transcription factor/ nuclear factor 1 (CTF/NF1)(2) and CCAAT binding protein (CBP)(3). To this end we synthesised oligonucleotide probes corresponding to the CCAAT motifs present in the rat albumin promoter, the HSV TK promoter and the high affinity NF1 binding site in the adenovirus inverted terminal repeat. In the first experiment (Figure 6, panel D) the rat albumin CCAAT competitor, but surprisingly, not the NF1 binding site, competed efficiently for CRF complex formation on probe IV. As this suggested that the factors recognising the EIIL CCAAT elements and the NF1 binding site were different, we next used a labelled NF1 binding site as probe and competed with bindings sites II, IV and VI (panel E). It was immediately obvious that the pattern of complexes formed using the NF1 probe is clearly distinct from that obtained using probes containing the EIIL CCAAT motifs (for a direct comparison compare lanes 1 and 5, panel F). Moreover, although the homologous NF1 binding site competes very efficiently for complex formation, no significant competition was observed using as competitors probes II and VI. In some experiments competition (2- to 3-fold) was observed using binding site IV (lane 3, panel E) and the HSV TK oligonucleotide (not shown) as competitors, but the affinity of NF1 for these sites was in the region of 50- to 100-fold lower than for the homologous probe, consistent with previous observations for the β-globin and TK CCAAT boxes (2). We conclude from these experiments that CRF and CTF/NF1 are distinct factors.

In contrast, the rat albumin promoter element competed very efficiently for CRF; indeed, using a rat albumin probe (probe Alb, Figure 4) essentially identical results to those described for EIIL region IV were obtained (panel G): that is, the EIIL CCAAT motifs compete for CRF binding but the NF1 binding site does not compete at all. The HSV TK element does, however, compete and use of the HSV TK CCAAT probe (panel H) confirms our observations that CRF can interact with a wide variety of CCAAT elements. In addition, the introduction of a C->G transversion into the first position of the TK CCAAT motif (probe TK*, Figure 4), a mutation known to severely reduce TK transcrip-

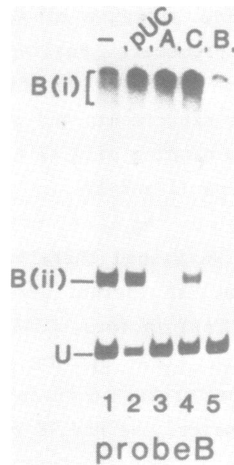

Figure 8. Competition for binding to probe B. Labelled probe B (see Figure 1B) was used in a gel electrophoresis DNA-binding assay. Competitor DNAs used at 50ng/assay are indicated.

tion in vitro and in vivo (3, 16), abolished CRF binding (panel I). Interestingly, this same mutation has been shown to enhance binding of CBP (3), suggesting that CBP is distinct from CRF, a conclusion supported by the observation that CBP is heat stable (17, 18), whereas CRF is heat sensitive (panel F).

Homologous recognition of EIIL binding site III and the c-fos enhancer

Our initial footprinting data showed a region of DNAse 1 protection covering the EIIL promoter from position -97 to -125 termed binding site III. An oligonucleotide probe corrresponding to this region did not compete for binding of the CRF (see Figure 6). To determine whether a cellular factor could recognise binding site III in the absence of other EIIL factor recognition sequences, we used a labelled oligonucleotide corresponding to sequences between positions -101 and -124 as a probe (probe III Figure 4) in the gel electrophoresis DNA-binding assay. The major complex formed (Figure 7a, lane 1) could be competed efficiently by the homologous DNA (lane 2) but not by either the rat albumin or binding site II oligonucleotides (lanes 4 and 5), and has a mobility distinct from that observed using the rat albumin oligonucleotide probe (compare lanes 6 and 7). A shortened oligonucleotide III^{s} (see Figure 4) also competed efficiently for the major retained band (lane 3) and could form a complex with similar mobility (not shown). We noted a strong homology between binding site III and an element juxtaposed to the serum regulatable element in the c-fos enhancer (Figure 7b). To test whether

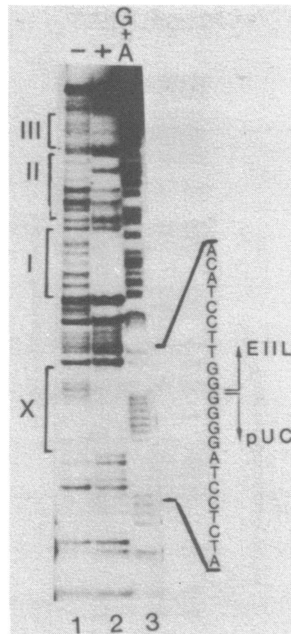

Figure 9. Footprinting across the pUC-EIIL junction. Footprinting reactions were as described for Figure 2 but using probe A 3′ end-labelled at the Hind III site. Footprints I-III correspond to those in figure 5 with protected region X covering the pUC-EIIL boundary.

this element could also bind the same factor, we used it as a probe and competitor in experiments similar to those described above. First, using the c-fos probe, a complex indistinguishable from that obtained using the EIIL binding site III probe was observed (Figure 7a, compare lanes 8 and 11). Second, both the c-fos and binding site III oligonucleotides cross-competed (lanes 10 and 12), strongly suggesting that both elements are recognised by the same factor. In a recent paper Piette and Yaniv (19) showed that a factor (termed PEA1) recognising this c-fos element could also footprint onto both the SV40 and polyoma enhancers at a region of sequence homology (see Figure 7b). To verify that the SV40 element bound the same factor as EIIL binding site III and c-fos, the SV40 sequences were synthesised and used as a probe and the complex formed in the gel electrophoresis-DNA binding assay compared with that obtained with EIIL probe III (lanes 14 and 15). To our surprise, the SV40 probe (AP1 in Figure 4, a gift from Dr H. Hurst) formed a complex with a markedly slower migration than that formed using EIIL probe III. In competition assays, the SV40 probe competed poorly for the complex formed on probe III and vice versa (H. Hurst, personal communication). Thus it appears

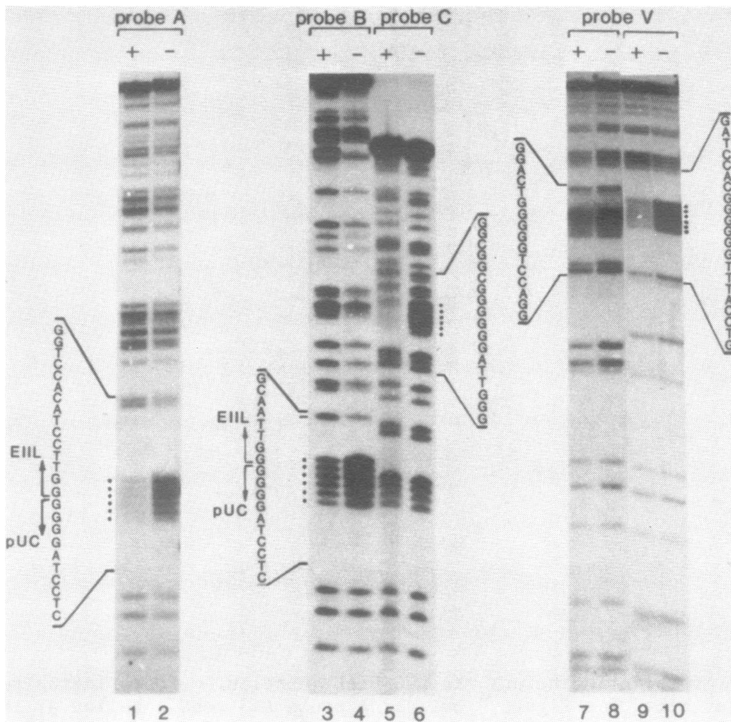

__Figure 10.__ DNA methylation interference. Probes A, B, C and V, 3′ end-labelled at the Hind III sites, were prepared and methylated __in vitro__ with DMS. Methylated DNA was then used in a gel electrophoresis DNA-binding assay. Retained DNA (+) (complexes equivalent to B(ii) in Figure 8) were excised and processed as described in the Materials and Methods. Unbound DNA (-) was similarly treated. Dots at the side of the sequence ladders indicate under-methylated G residues.

that the factors recognising the c-__fos__/EIIL elements and the SV40 ehancer sequences are different, despite the strong homology between binding sites. For convenience we refer to the factor recognising both binding sites III and the c-__fos__ element as __u__pstream __f__actor __II__ (USFII).

__Identification of a factor recognising the consensus sequence GGGGGGNT__

Cellular factors recognising binding sites I-IV and VI have been identified. However, no known consensus sequence was evident within binding site V. To attempt to characterise the cellular factors recognising this region we first used EIIL subclone B (Figure 1b) as a probe. Two major complexes were formed using this probe (Figure 8). Complex B(i) was unique to probe B and was not competed by probes A and C. In contrast, complex B(ii) was efficiently competed for both by the homologous probe and probe A, and to a

lesser extent by probe C. A complex with a similar migration to complex B(ii) was also observed using labelled probes A and C (not shown). Both complexes B(i) and B(ii) were formed using probe V (Figure 1b) corresponding to binding site V cloned into the pUC polylinker (not shown). Within binding site V a sequence GGGGGGNT was identified that was also present both adjacent to the CCAAT box in binding site VI (present on probe C) and at the pUC-EIIL junctions in probes A and B. To determine whether this sequence was recognised by a factor present in the nuclear extracts we attempted to obtain a footprint covering the pUC-EIIL junction in probe A. To this end an EcoRI - Hind III fragment correspondng to probe A and end-labelled at the Hind III site was prepared and digested with DNAse 1 in the absence or presence of nuclear extract (Figure 9, lanes 1 and 2 respectively). In addition to the protected regions observed previously using the intact EIIL promoter a footprint (x) was also obtained over the pUC-EIIL junction.

Further evidence for the interaction of cellular factors with the 'G_6' motifs came from a series of methylation interference experiments. Labelled fragments from probes A, B, C, and V were partially methylated in vitro with DMS prior to being used for the gel electrophoresis DNA-binding assay. The bands corresponding to complex B(ii) or equivalent complexes formed on probes A, C and V were excised, electroeluted and cleaved with piperidine as described (12) before being analysed on a sequencing-type gel. If methylation of any G residue interferes with complex formation then DNA methylated at that position will be underrepresented in the bound DNA compared to the unbound. Using a labelled fragment from probe A for such an experiment methylated G residues corresponding to the pUC-EIIL boundary and covered by the footprint shown in Figure 9 are clearly underrepresented in the complex compared to the unbound DNA (Figure 10, compare lanes 1 and 2). The G_6 motif at the pUC-EIIL boundary of probe B is similarly involved (lanes 3 and 4), as is the G_6 sequence adjacent to the CCAAT box in probe C (lanes 5 and 6), as well as the G_6 motif within binding site V and present in probe V (lanes 9 and 10). The stretch of five G residues present on the non-coding strand within binding site V does not appear to be involved (lanes 7 and 8). Thus, all four G_6 motifs examined bind a cellular factor and a run of five G residues is not sufficient for binding.

We have not yet defined precisely the sequences involved in the formation of complex B(i), which is therefore not discussed further.

DISCUSSION

The adenovirus EIIL promoter upstream region comprises a complex array of binding sites for cellular factors. A diagram summarising those identified in

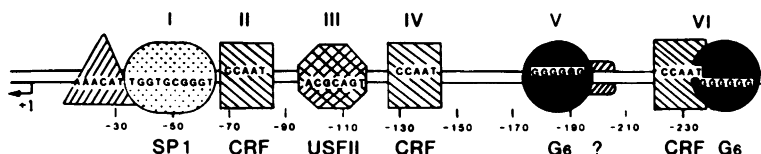

Figure 11. Summary of data presented. Schematic showing relative positions of binding sites and factors recognising them.

this study (by a combination of DNAse footprinting, gel electrophoresis-DNA binding assays, mutagenesis, competition and methylation interference experiments)is presented in Figure 11.

The capsite proximal binding site, region I, is located between positions -36 and -56 and available evidence suggests that this region is recognised by the cellular transcription factor Sp1. The factor recognising region I is efficiently competed for by the SV40 21bp repeats and at the heart of the DNAse 1 protected region lies a strong homology to other known Sp1 recognition sequences. Although this site does not conform to the full Sp1 consensus G/TGGGCGGPuPuP other similar non-consensus Sp1 binding sites have been described (13-15). These results are confirmed by the recent observations of Bhat et al (20), who demonstrated that affinity purified Sp1 footprinted efficiently within region I, and that linker-scanner mutations affecting the Sp1 consensus reduced EIIL promoter activity by up to 5-fold. In contrast, however, Sp1 bound very inefficiently to the GC-rich sequence located at position -80, within region II, supporting our data showing that these sequences did not compete for Sp1 binding to region I, and that they did not play a role in DNA-protein complex formation within region II. In this respect, these sequences are reminiscent of other apparent consensus Sp1 binding sites located in the promoters of the herpes simplex virus IE175 and Harvey ras genes which also fail to bind Sp1 in vitro (13, 14).

The identity of the factor interacting with binding sites II, IV and VI is more intriguing. Each of these sites contains a CCAAT motif and mutations affecting the C residues within this element eliminate binding. Two previously identified factors have been shown to bind such a sequence element. First, CTF was recently demonstrated to be homologous to a factor involved in the replication of the adenovirus genome, NF1 (2), and comprises a family of proteins with a molecular weight of between 55-65,000. Mutations within the CCAAT motif in the HSV TK promoter that reduce transcription efficiency similarly decrease binding by CTF/NF1. Moreover, addition of affinity purified CTF/NF1 to an in vitro transcription system increased transcription from the α-globin promoter (2). High affinity binding sites have been identified

within the adenovirus inverted terminal repetitions and in the α-globin and Ha-ras promoters, whereas the TK and β-globin CCAAT motifs are bound around 100-fold less efficiently (2). Second, a 20,000 molecular weight promoter protein termed enhancer binding protein (EBP20) which interacts with core elements within the SV40 and MSV LTR enhancers (18) is thought to be identical to a factor CBP previously shown to bind to the HSV TK and MSV LTR CCAAT boxes (3). CBP/EBP20 is a heat stable protein and like CTF/NF1, binding of this factor (to the TK and MSV CCAAT sequences) is reduced by mutations which decrease transcription from the MSV and TK promoters. However, introduction of C->G transversion in the first position of either the MSV or TK CCAAT motifs decreases transcription efficiency of the promoter but, in contrast with CTF/NF1, increases affinity for CBP/EBP20 (3). Several lines of evidence strongly suggest that the factor, CRF, recognising EIIL regions II,IV and VI is distinct from CTF/NF1 and CBP/EBP20. First, the high affinity NF1 binding site within the Ad-2 ITR does not compete for binding of CRF to any of the CCAAT elements tested, and similarly the CRF binding sites did not compete for NF1 binding. Second, the pattern of complexes formed by CRF and NF1 using the gel electrophoresis DNA- binding assay are very distinct and clearly different, and the pattern obtained for the NF1 probe is very similar to that obtained using affinity purified NF1 (2). Third, binding of affinity purified NF1 to probes containing EIIL binding sites II, IV and VI is not detected (L.Clark, R.Hay, personal communication). Fourth, preliminary experiments suggest that the molecular weight of CRF is greater than CTF/NF1 (our unpublished observations). Fifth, in contrast to CBP/EBP20, CRF is heat labile, and finally the C->G transversion, which increased binding of CBP/EBP20 to the TK and MSV CCAAT elements while decreasing transcription efficiency, abolished binding of CRF.

Although direct proof that CRF is intimately involved in transcriptional regulation will require its purification and use in reconstituted in vitro transcription systems, some evidence is available to suggest that it is. As mentioned above, the single point mutation in the TK CCAAT motif eliminates CRF binding and reduces TK transcriptional activity. The rat albumin CCAAT box, also recognised by CRF, is required for efficient transcription (21, 22) and in the EIIL promoter, linker-scanner mutations disrupting regions II and IV reduce EIIL transcription by up to 100- and 5-fold respectively (20). Deletion of the weak CRF recognition site at -230 does not affect EIIL transcription but it is widely accepted that 'upstream factors', as opposed to 'enhancer factors', have relatively stringent distance requirements. It is therefore possible that the -230 CCAAT motif is too far from the cap-site to exert an influence on EIIL transcription.

The identification of a factor, CRF, distinct from CTF/NF1 or CBP/EBP20, that recognises several functionally significant CCAAT elements is clearly of major importance. Although the CCAAT motifs are generally regarded as constitutive promoter elements, it is perfectly feasible that promoters might be regulated through these elements by virtue of the relative affinities of different CCAAT boxes for each of the cognate factors, CTF/NF1, CBP/EBP20 and CRF. This, combined with the relative abundance of each factor in an individual cell, as well as its ability to interact in a combinational fashion with other transcription factors, would provide a mechanism for differential regulation of genes sharing common promoter elements. Indeed, in the case of the Xenopus hsp70 gene the CCAAT box confers cell-type-specific regulation in oocytes (23). Further insights into the role of the CCAAT element will be obtained once reagents to the various CCAAT factors are available.

A different factor recognises the DNAase 1 protected region III centred on position -110. We identified a strong homology between the core of this sequence and an element in the c-fos enhancer juxtaposed to the serum regulatable element (see figure 8). The factor that bound this region, USFII, competed efficiently with the c-fos element and both sequences gave identical patterns in the band shift assay. While this work was in progress a report by Piette and Yaniv (19) identified a cellular factor termed PEA1 from mouse 3T6 cells that bound both to the same c-fos element and also to the polyoma and SV40 enhancers. It was suggested that this factor might be AP1, an SV40 enhancer binding protein recently purified to homogeneity by Lee et al (24). Although the homology between the binding sites for the EIIL region III and c-fos factors on the one hand, and the polyoma and SV40 enhancer elements on the other is great (see Figure 7b), we believe that they are recognised by distinct factors: the DNA-protein complex formed using the SV40 element in the gel electrophoresis-DNA binding assay gave a pattern clearly different from that obtained using either the c-fos or EIIL region III sequences, and cross-competition experiments (H.Hurst, personal communication) gave results consistent with the existence of two factors with related but distinguishable sequence requirements.

Analysis of both the adenovirus early region EIIa and EIV promoters has led to the identification of cellular factors with sequence requirements for binding apparently identical to those of USFII (25-27). A similar factor has also been described as recognising an homologous sequence in the c-fos promoter at position -67. In all cases, EIIL (20), EIIAE (25, 26), EIV (27) and c-fos (28), the elements recognised are required for efficient transcription. Indeed, duplication of the EIV element mediates constitutive enhancer function in vitro.

Thus so far three different upstream factors recognising EIIL binding sites I-IV appear to be implicated in EIIL promoter function. According to Bhat et al, sequences upstream of region IV are dispensible for efficient EIIL expression. Thus, the role of the two G_6 elements at positions -197 and -234 remains unclear. Of the four sites identified that bind the G_6 factor (two occurring at the pUC-EIIL junctions in EIIL subclones) the only obvious consensus for binding is GGGGGGNT. Interestingly, this motif is represented in the Ad-2 major late promoter adjacent to the TATA box, and a recent report (29) suggested that it may be required for efficient MLP function. Clearly, more work is required to elucidate the role, if any, of the G_6 binding protein in the mechanism of eukaryotic transcription.

During adenovirus infection, expression from the EIIL promoter is very weak early in infection and very high late, after the onset of DNA replication. What possible mechanisms for this switch can be envisaged? It is possible that products of the EIa or other early regions may play a role. According to one report, the product of the 12S EIa message represses EIIL expression (6). However, other workers indicate that EIa products do not affect EIIL promoter activity (5)(B.Thimmappaya, personal communication). The contrast with the EIa inducible early promoters is striking, all the more so since EIIL may share the factor recognising binding site III with the EIa inducible EIIa-early and EIV promoters. It may be that the presence of a single such factor recognition site is insufficient to confer EIa inducibility and that a dimer (as in EIV), or cooperation with other factors (as in EIIaE), may be required. In the case of EIIa-early, a cellular factor is induced by EIa and is thought to be involved in EIa-mediated induction of that promoter (30). We have examined extracts made from both early- and late-infected HeLa cells, but have failed to show any differences between these extracts in terms of abundance of factors recognising EIIL (unpublished observations). It is possible, however, that the effective concentration of a particular EIIL transcription factor is increased late in infection, perhaps by redistribution of sequestered factors, but that preparation of the nuclear extracts masks any such differences. Alternatively, enhanced EIIL expression may simply be elevated late in infection as a result of the increased copy number of replicated viral templates. Whichever of these possibilities is correct, it is clear that EIIL provides an interesting system for studying the molecular mechanisms of regulation of a complex eukaryotic promoter.

ACKNOWLEDGMENTS

We are indebted to Dr.P.Whiteside and the Department of Microbiology at the University of Surrey for their help and support in the large scale culture

of HeLa cells. We would like to thank Drs.R.Hay, L.Clark and G.Currie for useful discussions, Drs.T.Leff and R.Hay for the gifts of recombinants pLi and NF1 respectively and Dr.P.O'Hare for critically reading this manuscript. We also wish to acknowledge Dr.B.Thimmappaya for communicating results prior to publication and Mrs.J.Mollan for preparing the manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Briggs,M.R., Kadonaga,J.T., Bell,S.P. and Tjian,R.(1986)Science **234**, 47-52
2. Jones,K.A., Kadonaga,J.T., Rosenfeld,P.J., Kelly, T.J. and Tjian,R.(1987) Cell **48**, 79-89
3. Graves,B.J., Johnson,P.F. and McKnight,S.L.(1986) Cell **44**, 565-576
4. Baker,C. and Ziff,E.(1981) J. Mol. Biol. **149**, 189-221
5. Leff,T. and Chambon,P.(1986) Mol. Cell. Biol. **6**, 201-208
6. Guilfoyle,R.A., Osheroff, W.P. and Rossini,M.(1985) EMBO J. **4**, 707-713
7. Galas,D. and Schmitz,A.(1978) Nucleic Acids Res. **5**, 3157-3170
8. Fried,M. and Crothers,D.M.(1981) Nucleic Acids Res. **9**, 6505-6525
9. Garner,M.M. and Revzin,A.(1981) Nucleic Acids Res. **9**, 3047-3060
10. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G.(1983) Nucleic Acids Res. **11**, 1475-1489
11. Bradford,M.M.(1976) Anal. Biochem. **72**, 248-254
12. Maxam,M.A. and Gilbert,W.(1980) Meth. Enzymol. **65**, 499-560
13. Ishii,S., Kadonaga,J.T., Tjian,R., Brady,J.N., Merlino,G.T. and Pastan,I. (1986) Science **232**, 1410-1413
14. Jones,K.A. and Tjian,R.(1985) Nature **317**, 179-182
15. Jones,K.A., Kadonaga,J.T., Luciw,P.A. and Tjian,R.(1986) Science **232**, 755-759
16. Jones,K.A., Yamamoto,K.R. and Tjian,R.(1985) Cell **42**, 559-572
17. McKnight,S.L. and Tjian,R.(1986) Cell **46**, 795-805
18. Johnson,P.F., Landschulz,W.H., Graves, B.J. and McKnight,S.L.(1987) Genes and Devel. **1**, 133-146
19. Piette,J. and Yaniv,M.(1987) EMBO J. **6**, 1331-1337
20. Bhat,G., SivaRaman,L., Murthy,S., Domer,P. and Thimmappaya,B.(1987) EMBO J. **6**, 2045-2052
21. Heard,J.M., Herbomel,P., Ott,M.-O., Mottura-Rollier,A., Weiss,M. and Yaniv,M.(1987) Mol. Cell. Biol. **7**, 2425-2434
22. Gorski,K., Carneiro,M. and Schibler,U.(1986) Cell **47**, 767-776
23. Bienz,M.(1986) Cell **46**, 1037-1042
24. Lee,W., Mitchell,P. and Tjian,R.(1987) Cell **49**, 741-752
25. SivaRaman,L., Subramanian,S. and Thimmappaya,B.(1986) Proc. Natl. Acad. Sci. USA **83**, 5914-5918
26. Boeuf,H., Zajchowski,D.A., Tamura,T., Hauss,C. and Kedinger,C.(1987) Nucleic Acids Res. **15**, 509-528
27. Lee,K.A.W. and Green,M.R.(1987) EMBO J. **6**, 1345-1353
28. Gilman,M.Z., Wilson,R.N. and Weinberg,R.A.(1986) Mol. Cell. Biol. **6**, 4305-4316
29. Brunet,L.J., Babiss,L.E., Young,C.S.H. and Mills,D.R.(1987) Mol. Cell. Biol. **7**, 1091-1100
30. Kovesdi,I., Reichel,R. and Nevins,J.R.(1986) Cell **45**, 219-228.