Volume 17 Number 23 1989 Nucleic Acids Research

Binding sites of HeLa cell nuclear proteins on the upstream region of adenovirus type 5 E1A gene

Koichi Yoshida, Mitsuo Narita and Kei Fujinaga

Department of Molecular Biology, Cancer Research Institute, Sapporo Medical College, S-1, W-17, Chuo-ku, Sapporo 060, Japan

Received July 17, 1989; Revised and Accepted October 27, 1989

ABSTRACT

Twenty one binding sites of HeLa cell nuclear proteins were identified on the upstream region of adenovirus type 5 E1A gene using DNase I footprint assay. The proximal promoter region contained five binding sites that overlapped the cap site, TATA box, TATA-like sequence, CCAAT box, and -100 region relative to the E1A cap site(+1). The -190 region was a potential site for octamer-motif binding proteins, such as NFIII and OBP100. An upstream copy of the E1A enhancer element 1 was the site for a factor (E1A-F) with the binding specificity of XGGAYGT(X=A, C; Y=A, T). E1A-F factor also bound to three other sites, one of which coincided with the distal E1A enhancer element. The distal element also contained a potential site for ATF factor. The adenovirus minimal origin of DNA replication competed for DNA-protein complex formation on the CCAAT and TATA box region and the -190 region, suggesting that these regions interacted with a common or related factor.

INTRODUCTION

The adenovirus(Ad) E1A gene was shown to be the Ad transcription unit first activated in a productive infection (1) and the E1A product stimulated transcriptions from other early gene regions on the viral genome (for review, see ref. 2). Relatively few studies have been performed on transcription control of the E1A gene itself. An upstream region of the adenovirus type 5 (Ad5) E1A gene extends from the left-terminal end of the viral genome (nucleotide 1) to the E1A cap site (nucleotide 499)(3). This region contains the replication origin of Ad DNA and transcriptional regulatory elements of the E1A and other early genes. Deletion mutagenesis defined a distal transcription enhancer element, located from positions -322 to -343 relative to the E1A cap site (+1)(4). Additional enhancer elements have been located using viral mutants with small, random deletions (5,

6). The element 1 specially regulated E1A transcription. The element II, located between repeats of the element 1, regulated transcription *in cis* of all other early genes on the viral genome in infected cells. Next to both repeats of the element 1, there were binding sites for the HeLa cell factor, E2F, which was initially found to be important for E1A-induced transcriptional activation of the early region E2A promoter (7). In the proximal promoter region, TATA and CCAAT boxes are evident (8). DNA sequences surrounding the TATA box, CCAAT box, and the cap site are highly conserved among human adenoviruses (8). Deletion studies demonstrated that the TATA box was required for precise initiation of transcription *in vitro* and *in vivo* and the CCAAT box had relatively little effects on the E1A transcription (9, 10). It is likely that these *cis*-acting regions may be ocuppied by binding with specific cellular transcription factors. A study of *cis*-acting regulatory sequences and the protein which interacts with them is essential to understand the mechanism by which the E1A gene is regulated at a transcription level.

Here, we have identified twenty one binding sites of HeLa cell nuclear proteins on the upstream region of Ad5 E1A gene by DNase I footprint assay. Comparisons of these binding proteins with each other were performed by methylation interference and competition binding assays.

MATERIALS AND METHODS

Cells and extracts

HeLa cell SIII line was grown in suspension in Dulbecco's modified MEM with 5% horse serum and harvested at 4-6x10^5 cells/ml. Nuclear extracts were prepared by the method of Dignam et al. (11) as modifed by Barrett et al. (12). Nuclear extract was fractionated by a chromatography on heparin-agarose (Pharmacia) as described by Lichtsteiner et al. (13). Briefly, the proteins bound to heparin-agarose were collected by stepwise elution using 2.5 column volumns of buffer A without KCl (13) but containing 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 1.5M NaCl. Protein concentrations were determined by the method of Bradford using Bio Rad assay kit.

Probes

Probe DNAs were derived from subclones of the E1A upstream sequences. The BamHI(-498, relative to the E1A cap

site, with a BamHI linker)/PvuII (-45) fragment was isolated from p5XhoC plasmid(14) and cloned at the BamHI site of pUC13 by converting the PvuII to BamHI sites with a linker (pUC -498/45). The BamHI insert of pUC-498/45 was cut with restriction enzymes into smaller fragments : BamHI(-498, with a BamHI linker)/AccII(-427); AccII(-426)/RsaI(-305); RsaI(-304)/SspI(-158); SspI(-157)/BamHI(-45, with a BamHI linker). These fragments were converted into BamHI fragments by linkers and cloned into the BamHI site of pUC13. The SspI(-157)/BglII(+50, with a BglII linker) fragment was isolated from p5SB16 plasmid (Ishino and Fujinaga, unpublished) and cloned between the HincII and BglII sites of the pUC13 carrying BglII linker at the HindIII site. DNA was labeled at the 5' -end using T4 polynucleotide kinase and (γ-^{32}P)ATP(NEN, sp. act. 3,000 Ci/mmol).

Gel retardation assay

Standard binding reactions containing nuclear extracts and 5' -end labeled probes were carried out in 25 µl mixtures in binding buffer : 20 mM Hepes-NaOH PH 7.8, 50 mM NaCl, 0.5 mM DTT, 0.5 mM EDTA, 5% glycerol, 2 µg of poly dI-dC/dI-dC(Pharmacia P-L Biochemicals), and 5 µg of crude nuclear extracts or 3 µg of heparin-agarose fractions. For competitions, 30 or 200 fold molar excess DNA was added into the mixtures. Binding mixtures were kept on ice for 10 min and then ^{32}P-labeled probe DNA (0.1 to 0.5 ng/reaction) was added. After incubation for 15 min at room temperature, reaction mixtures were resolved in 4% non-denaturing polyacrylamide gel (acrylamide:bis-acrylamide, 40:1).

DNase I footprint assay

Binding mixtures of 50 µl contained gel retardation binding buffer, 0.2 -0.6 ng of ^{32}P-labeled probe DNA, and 40 µg of crude extracts or 20 µg of heparin-agarose fractions or BSA(BRL). Binding mixtures were incubated as described for gel retardation assays and then 5 µl of DNase I(Worthington Diagnostics, 1 mg/ml stock solution was freshly diluted into the buffer of 10 mM Tris-HCl pH 7.5 and 50 mM $MgCl_2$ at 100 µg/ml) was added and kept for 1 min at room temperature. DNase I digestions were stopped as described (15). The region protected from DNase I weas resolved by electrophoresis through 8% or 12% polyacrylamide gels containing 8M urea with A+G marker (16).

Methylation interference assay

Singly end-labeled DNA (5 x 10^5 cpm) was partially methylated by treating for 3 min with 1 µl Dimethyl sulfate(DMS) as described (16). After ethanol-precipitation twice, DNA was incubated with 20 µg of heparin-agarose fraction in a mixture scaled up 4-fold from analytical binding reactions and electrophoresed in 4% non-denaturing polyacrylamide gel. After autoradiography, DNA was eluted from the region of gels containing free DNA and protein-DNA complexes as described by Treisman (17). DNA was purified from co-eluted acrylamide by a NACS column chromatography(BRL) according to supplyer's mannual. DNA was cleaved with 100 µl of 1M piperdine and equal amonts of DNA (in Cerenkov) were electrophoresed in 8% or 12% polyacrylamide gels containing 8M urea with A + G Marker (16).

Oligonucleotides

Complementary single-stranded synthetic oligonucleotides containing Ad5 nucleotide sequences 1 to 20 (BS21 oligo) and nucleotides 300 to 320(BS7 oligo) were made using a Milligen DNA synthesizer Model 7500 and partially purified by NEN sorb columns (NEN Dupont) according to supplier's mannual. Complementary oligonucleotides were annealed by first heating to 100 C in 0.3M NaCl, 10 mM Tris-HCl pH8.0 and 1 mM EDTA, followed by slowly cooling.

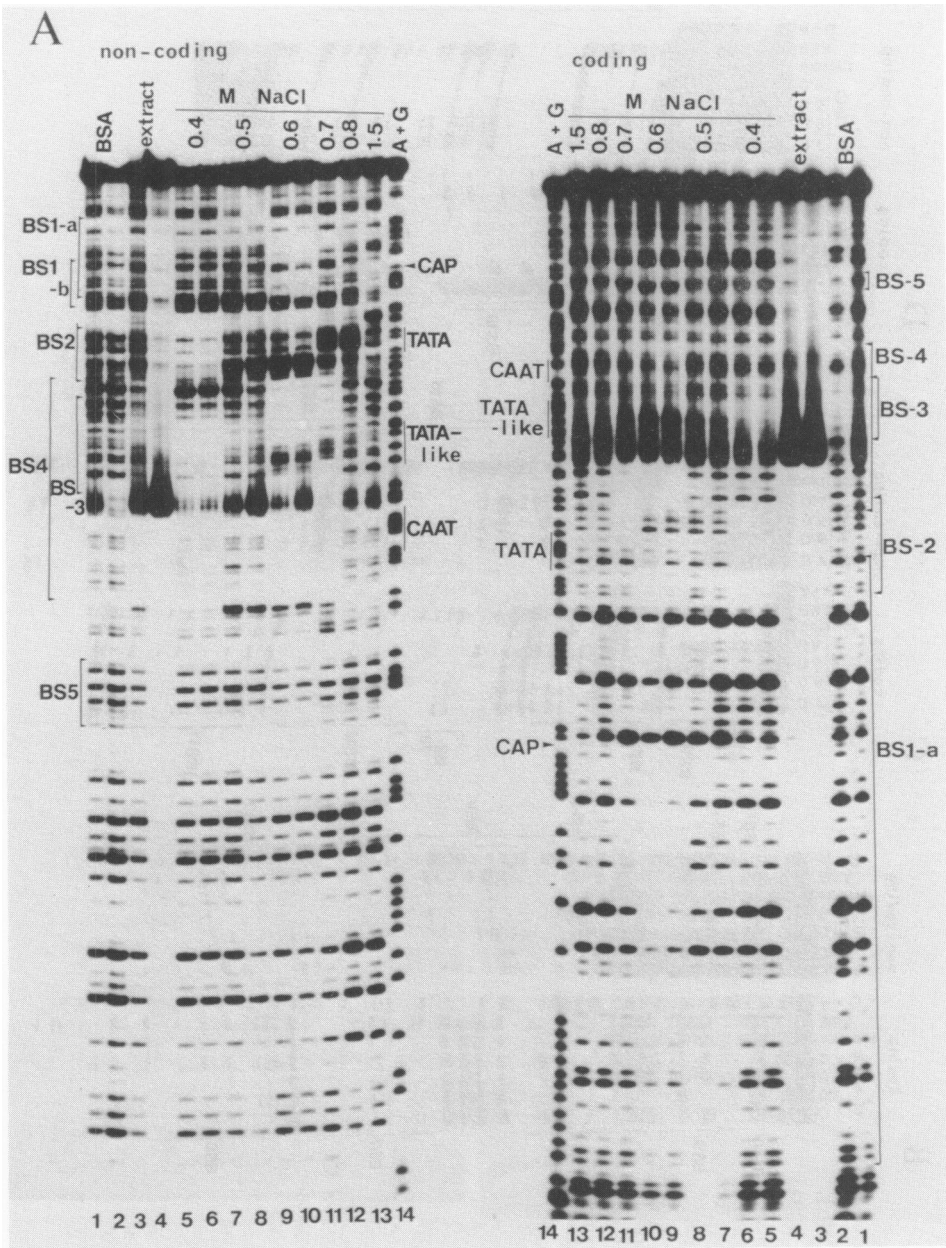
RESULTS

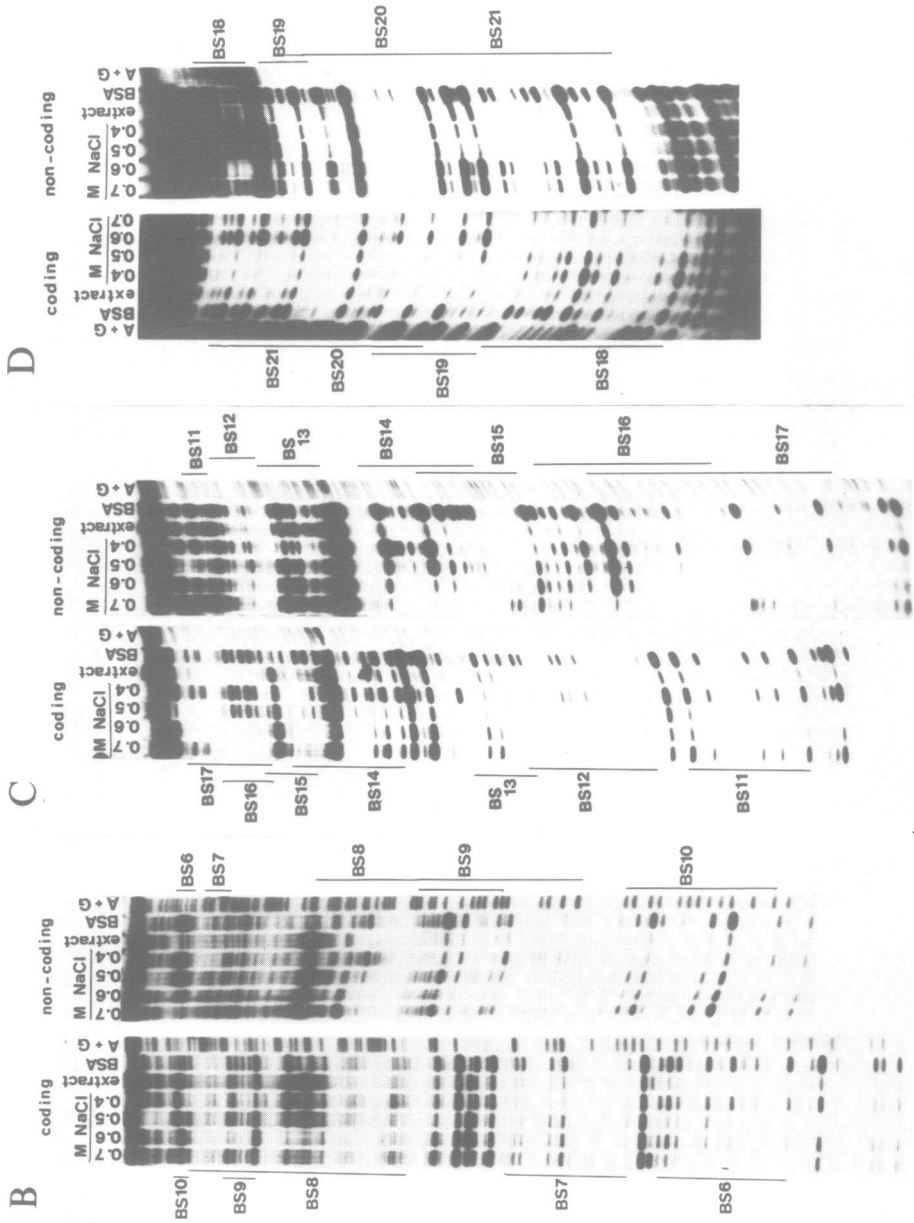
Binding sites of cellular proteins on the E1A promoter, E1A enhancer, and replication origin

We analyzed proteins that interacted with the 5' upstream region of the Ad5 E1A gene using DNase I footprint assay (18). To improve resolution of protein/DNA complexes and determine precise boundaries of individual binding sites, we fractionated crude extracts by a chromatography on heparin-agarose using stepwise elution method. Each fraction was incubated with 5' end-labeled DNA probes and digested with DNase I. Figure 1 shows the presence of at least twenty one binding sites as revealed by protections from DNase I. The locations of these binding sites were as follows.

(1) Probe C labeled at position +50 on the coding strand and labeled at position -157 on the noncoding strand revealed cfive protected regions (Fig. 1A, BS1 to 5). The precise

boundaries of the protected regions were shown in Figure 2. BS1, which includes the E1A cap site contained two overlapping binding sites of BS1-a and b due to binding activities eluted




D
BS18
BS19
BS20
BS21
non-coding
coding
M NaCl
0.4
0.5
0.6
0.7
extract
BSA
A + G
BS21
BS20
BS19
BS18
C
BS11
BS12
BS 13
BS14
BS15
BS16
BS17
B
BS6
BS7
BS8
BS9
BS10

in 0.6M and 1.5M NaCl fractions, respectively (Fig. 1A). The footprint at the BS1-a extended from -25 to +34 and was seen using the coding and noncoding strands(Fig. 1A, lanes 9 and 10). The footprint at the BS1-b extended from -12 to +9 and was detectable with only a noncoding strand(Fig. 1A, lane 13). However, methylation interference assay demonstrated that both strands interacted with the protein (Fig. 6). BS2 was between -21 and -38 and coincided with the TATA box(TATTTATA) of the E1A gene (Fig. 2). The footprint at BS2 was seen using the 0.4M NaCl fraction (Fig. 1A, lanes 5 and 6). BS3 was between -45 and -72 and contained TATA-like sequence (TATTATTATA : -59 to -50, Fig. 2), protection of which was seen using the 0.4M NaCl fraction (Fig. 1A, lanes 5 and 6). BS4 was between -36 and -83 and overlapped the BS3 site. It contained the CCAAT box (GTCAAAGT, Fig. 2). The footprint at BS4 was seen using 0.6M and 0.7M NaCl fractions, but the profiles differed clearly (Fig. 1A, lanes 9, 10, and 11). BS4 may contain more than one binding site. The CCAAT and TATA box region were also protected with the 0.4M NaCl fraction using the noncoding strand (Fig. 1A, lanes 5 and 6). BS5 extended from -99 to -91 and the footprint was seen using 0.6M and 0.7M NaCl fractions but it was very faint (Fig. 1A, lanes 9 and 10). The BS5

Figure 1. DNase I footprinting on the upstream region of Ad5 E1A gene. Binding reactions of the indicated probes with BSA, crude HeLa cell nuclear extracts (extract), or heparin-agarose fractions were carried out (see methods). The top of lanes indicates the molar NaCl concentration at which the protein used in the assay was step-eluted. A+G are A+G sequencing lanes(16). The footprints (BS1 to 21) are indicated by vertical lines. A. Footprinting on the proximal promoter region. Probe C labeled at +50 on the noncoding strand or -157 on the coding strand was incubated with the indicated fractions containing following amounts of proteins: Lanes 1 and 2, 20 μg ; Lane 3, 40 μg ; Lane 4, 80 μg ; Lanes 5, 7, 9, 11, 12 and 13, 20 μg ; Lanes 6, 8 and 10, 40 μg. The positions of the cap site, TATA box, TATA-like sequence and CCAAT box are indicated. B. Footprinting on the enhancer region. Probe B labeled at -158 on the coding strand or -304 on the noncoding strand was incubated with 20 ug of BSA, 40 μg of crude nuclear extracts, and 20 μg each of heparin-agarose fractions. C. Footprinting on the distal enhancer region using probe A labeled at -305 on the coding strand or -426 on the noncoding strand. Added amounts of proteins are described in panel B. D. Footrpinting on the viral DNA replication origin using probe ori labeled at -427 on the coding strand or -498 (the left end of viral DNA) on the noncoding strand. Addd amounts of proteins are described in panel B.

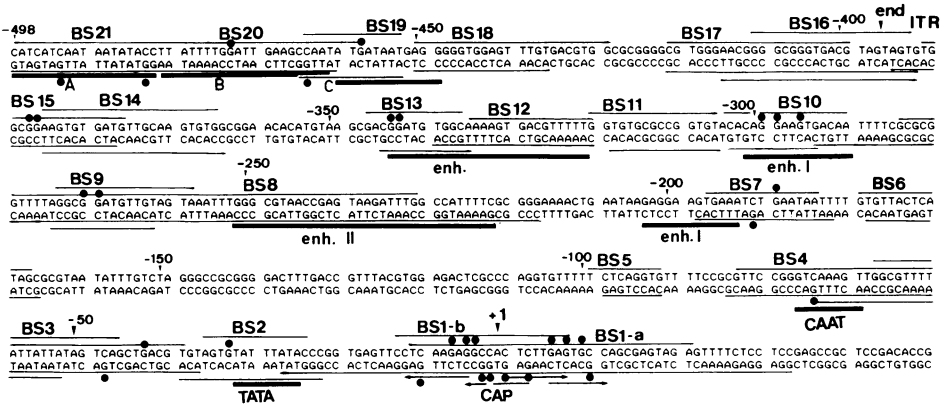


Figure 2. Summary of DNase I footprints on the upstream sequence of Ad5 E1A gene. The thin horizontal lines indicate the regions of DNase I protection (BS1 to 21). The numbers are relative positions to the E1A cap site (+1, 498 bp upstream from the left-end of Ad5 DNA). The bars indicate the positions of the TATA and CCAAT boxes, enhancer elements I and II, distal enhancer, and three domains (A, B, and C) constituting the Ad5 origin of DNA replication(34). Arrows denote hyphenated symmetry at the cap site. Dots indicate the G residues whose methylation interfere with protein-bindings.

footprint was improved using the probe labeled at the PvuII site at position -45 (data not shown). When incubated with the 0.5M NaCl fraction, probe DNAs were often protected at both ends from DNase I(Fig. 1A, lanes 7 and 8; Fig. 1B and 1C). It might be due to a non-specific binding of proteins to DNA termini.

(2) Probe B labeled at position -158 on the coding strand or position -304 on the noncoding strand revealed five footprints (Fig. 1B, BS6 to 10). BS6 extended from -165 to -178, protection of which was seen using the 0.7M NaCl fraction. BS7 was between -181 and -197. The footprint was detected with the 0.4M NaCl fraction and partially overlapped a downstream copy of the enhancer element 1 (6)(Fig. 2). BS8 extended from -221 to -293 and the foorprint was observed with the 0.6M NaCl fraction. BS8 contained the enhancer element II (6)(Fig. 2) and contained two binding sites for the E2F factor (TTTTCGCG: -218 to -225 and -281 to -288, Fig. 2) which appeared to be involved in E1A-induced transcriptions of the E2 gene and E1A gene itself (7). BS9 was between -257 and

-273 and the footprint was seen using the 0.4M NaCl fraction. BS10 was between -289 and -304 and coincided with an upstream copy of the E1A enhancer element I (6)(Fig. 2). The BS10 footprint was observed with the 0.4M NaCl fraction.

(3) Probe A labeled at position -305 on the coding strand or position -426 on the noncoding strand revealed seven footprints (Fig. 1C, BS11 to 17). BS11 and 12 extended from -305 to -319 and -319 to -338, respectively (Fig. 2). The footprint at the BS11 was observed with the 0.6M and 0.5M NaCl fractions. The footprint of BS12 was seen using 0.6M and 0.7M NaCl fractions. BS13 was between -334 and -346, footprint of which was detected with the 0.4M NaCl fraction. BS12 and 13 overlapped the distal E1A enhancer element that has been localized from -320 to -343 (Fig. 2) by transient expression assay using the E1A upstream DNA with the various extents of deletions (4). BS14 extended from -363 to -384 and the footprint was seen using 0.5M, 0.6M, and 0.7M NaCl fractions. BS15 was between -375 and -394 and the footprint was observed with the 0.4M NaCl fraction. BS16 extended from -391 to -414 and the footprint was detectable with the 0.7M NaCl fraction. BS17 was between -391 and -420 and the footprint was seen using the 0.6M NaCl fraction.

(4) Probe ori labeled at position -427 on the coding strand or position -498 on the noncoding strand showed four footpirnts (Fig. 1D, BS18 to 21). BS18 was between -450 and -428, footprint of which was observed with the 0.6M NaCl fraction. BS19 extended from -452 to -464 and the footprint was found in the 0.4M NaCl fraction. The BS19 overlapped domain C, a binding site for the NFIII (or ORP-C) factor (19, 20)(Fig. 2). The footprint extending from -458 to -498 (the left end of viral DNA) was detectable with 0.4M and 0.5M NaCl fractions. This footprint overlapped domains B and A (Fig. 2), binding sites for the NF1 and origin recognition protein A (ORP-A), respectively (21-25). We could not see a boundary between these sites. To describe clearly, we refered to the potential binding sites of the NF1 and ORP-A as BS20 and 21, respectively.

Ad minimal origin of DNA replication competes for bindings to the proximal E1A promoter region

Ad minimal origin (nucleotides 1 to 18, domain A in Fig. 2) contains a AT strech(TATATTAT, nucleotides 16 to 9)

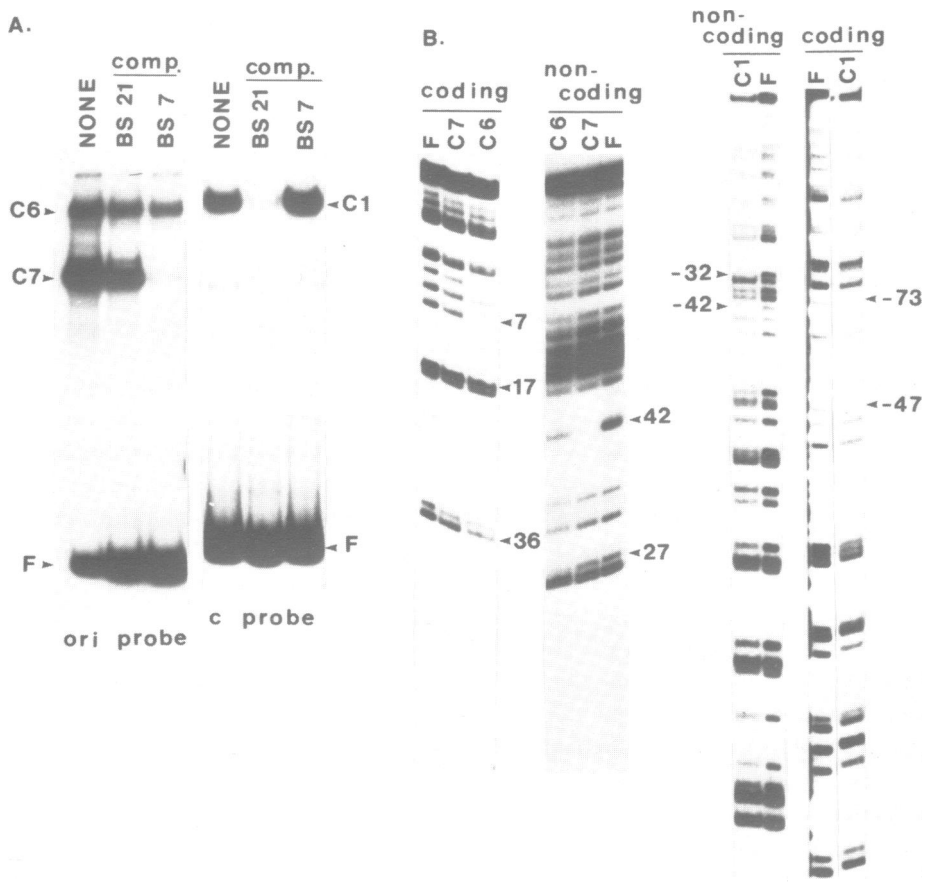


Figure 3. Ad minimal origin competes for bindings to the proximal promoter region containing the E1A CCAAT and TATA boxes. A. Gel retardation assay of the Ori(-427 to -498) and C(+50 to -157) probes with the 0.4M NaCl heparin-agarose fraction. For competitions, 200-fold molar excess of double stranded oligonucleotide was addd to each incubation, as indicated. BS21, Binding site 21 oligomer (20 bp, -479 to -498); BS7, Binding site 7 oligomer (20 bp, positions -199 to -180); C1, C6, and C7, gel retardation complexes; F, Free probe DNA. B. Methylation interference analysis of the complexes, C1, C6, and C7. Partially methylated DNA of probes ori and C were used in a preparative gel retardation assay. Free (F) and complexed DNA(C1, C6, and C7, panel A) were recovered from gels, cleaved with piperdine, electrophoresed in 8% or 12% polyacrylamide sequencing gels. The G residues whose methylation interfere with complex formation are marked by relative positions to the E1A cap site.

conserved perfectly in human adenoviruses (8). It is similar to the TATA box. We examined whether the minimal origin could compete for bindings to the E1A TATA box region. DNA oligonucleotides of 20 bp (BS21 oligo) homologous to the minimal origin were synthesized. A and T-rich DNA oligonucleotides of 20 bp(BS7 oligo) homologous to the BS7 site but different from BS21 oligo was used as a negative control. Origin probe containing the minimal origin was incubated with the 0.4M NaCl fraction of heparin-agarose and resulting complexes were resolved by gel retardation assay (26)(Fig. 3A). Two complexes of C6 and C7 were detected. Methylation interference assay indicated complex 6 resulted from bindings of the protein to the minimal origin and the NF1-site (Fig. 3B). G residues at nucleotides 27 (-472 relative to the E1A cap site) on the noncoding strand and 36(-463) on the coding strand were in close contacts with the protein, probably NF1, although the purified NF1 revealed three more contactpoints with the protein (27). Two G residues at 7 (-492) and 17(-482) on the coding strand within the minimal origin were also in close contacts with the protein, probably, ORP-A (21)(Fig. 3B). Complex 7 resulted from bindings of the NFIII factor to the BS19, because methylation interference profile (G at nucleotide 42, Fig. 3B) was similar to that of purified NFIII(19, 20). Self-competition with the BS21 oligo homologous to the minimal origin slightly decreased a level of complex 6(Fig. 3A). Complex 6 might be stabilized by protein-protein interaction of ORP-A and NF1 factors. Complex 6 also contained NFIII factor, because weak interference of G at nucleotide 42 on the NFIII site was detectable in complex 6 (Fig. 3B). Complex 7 disappeared upon addition of the excess BS7 oligo(Fig. 3A), indicating that the BS7 site may bind to the NFIII factor(described later). Using probe C (-157 to +50), a single complex (C1) was detected with the 0.4M NaCl fraction of heparin-agarose (Fig. 3A). Complex 1 resulted from bindings of the protein to the E1A TATA and CCAAT box region. There were two G residues in close contacts with the protein on the noncoding strand at position -32, next to the first T of the E1A TATA box and at position -42 (Fig. 3B). On the coding strand, two G residues at position -73 on the CCAAT box

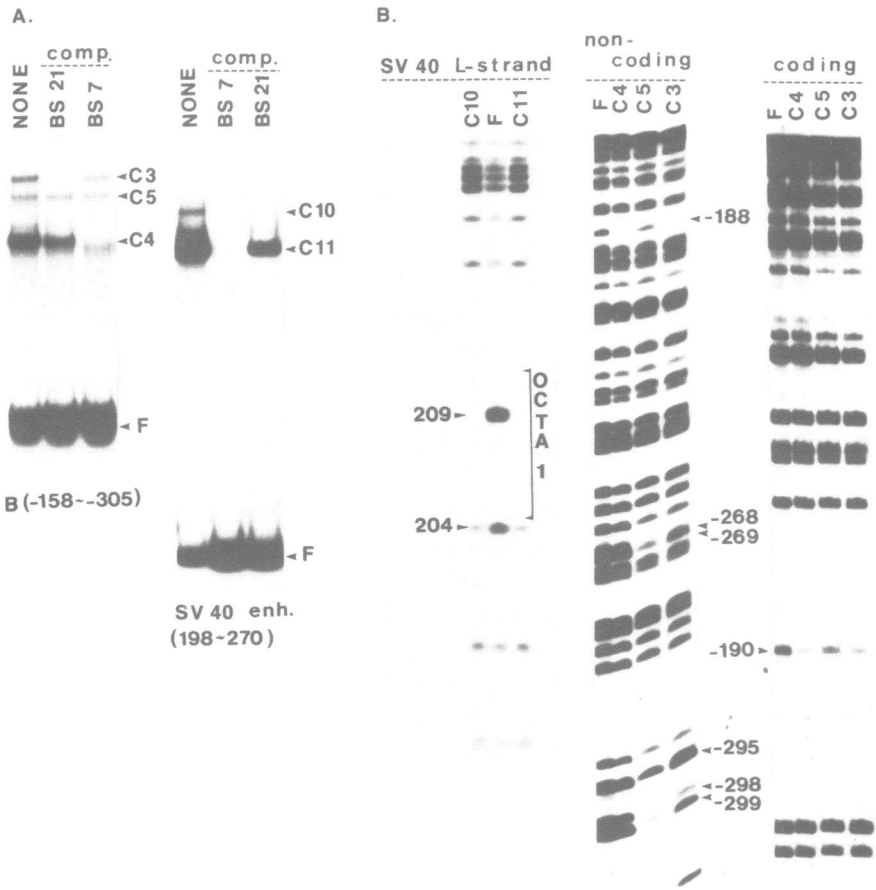


Figure 4. The oligonucleotide of BS7 site competes for bindings to the Ad NFIII and SV40 octa-1 binding sites. A. Gel retardation aasays of the probe B (-158 to -305) and SV40 enhancer fragment (PvuII/NsiI, 198 to 270) by the 0.4M NaCl fraction of heparin-agarose. Gel retardation and competition assays were performed as described in Figure 3. B. Methylation interference analysis of the gel retardation complexes C3, C4, C5, C10, and C11. The G residues whose methylation interfere with complex formation are indicated in SV40 sequence positions(C10, C11) and in relative positions to the Ad5 E1A cap site(C3, C4, C5). The position of the octamer-related SV40 sequence octa-1 is shown.

and at position -47 were in close contacts with the protein (Fig. 3B). The result was consistent with the DNase I footprint data that the TATA and CCAAT box region was

protected by incubation with the 0.4M NaCl fraction (Fig. 1A, lanes 5 and 6). Competition with the excess BS21 oligo decreased a level of complex 1, whereas the BS7 oligo had no effects (Fig. 3A). The result suggested that the minimal origin and proximal E1A promoter interacted with a common factor.

A factor bound to the -190 region is similar to the Ad and SV40 octamer-motif binding HeLa cell factor.

Previous competition binding assay (Fig. 3A) demonstrated that the -190 region(BS7 site) may interact with the octamer-motif binding factor, NFIII. Probe B that contains the -190 region produced three gel retardation complexes 3, 4, and 5 by incubation with the 0.4M NaCl fraction of heparin agarose (Fig. 4A). Both of complexes 3 and 4 represented protein-bindings to the -190 region. No significant differences in methylation interference patterns were found between complexes 3 and 4: one G residue was in close contacts with the protein at -188 on the noncoding strand and at -190 on the coding strand (Fig. 4B). Self-competition with the excess amount of BS7 oligo abolished complex 4, while a level of complex 3 reduced slightly (Fig. 4A). Rather, complex 3 was decreased by competition with the minimal origin sequence (BS21 oligo), suggesting the presence of minimal origin-binding factor in complex 3. Complex 5 remained intact because of bindings to the BS9 and 10 sites (see Fig. 5). Next, the SV40 PvuII/Nsi fragment (positions 198 to 270) that contains the octa-1 of the two adjacent octamer-related sequences(octa-1 and -2) of the SV40 enhancer element B was used as a probe. One major and three minor complexes were generated with the 0.4M NaCl fraction of heparin agarose (Fig. 4A). Addition of the BS21 oligo had a marginal effect on complexes, whereas the BS7 oligo abolished complex formation (Fig. 4A). Methylation inteference analysis of complexes 10 and 11 exhibited a common interference pattern, representing protein-binding to the octa-1 site of SV40 enhancer element B (28): G resiudes at 209 and 204 were in close contacts with the protein on the SV40 late strand (Fig. 4B). Taken together, we concluded that the factor bound to the -190 region was identical or related to adenovirus and SV40 octamer-binding HeLa cell proteins, NFIII (or ORP-C) and OBP100.

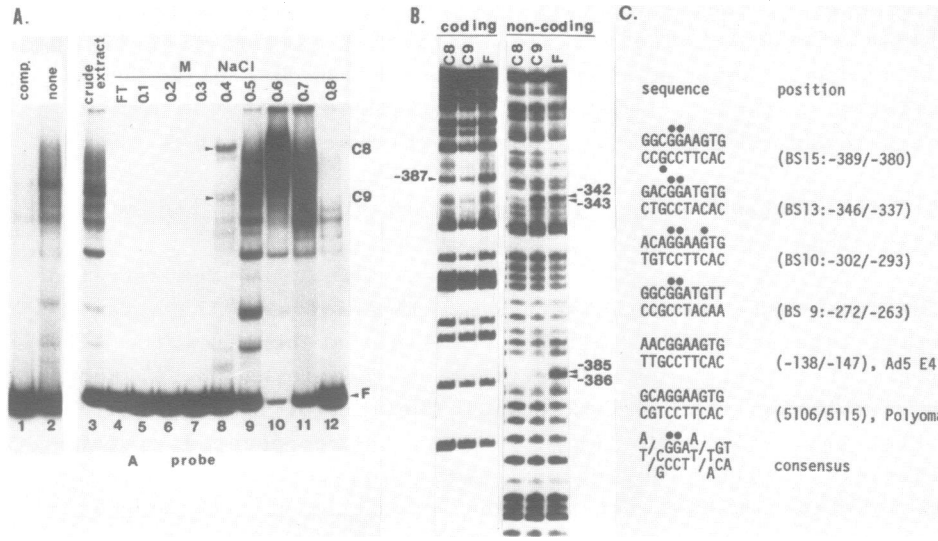


Figure 5. Four different sites of BS9, 10, 13, and 15 interact with a common factor E1A-F. A. Gel retardation assay of A probe containing the distal element of E1A enhancer. Probe A (-305 to -426) was incubated with 5 μg protein of crude extract (lanes 1 to 3) and 3 μg protein of each fraction of heparin-agarose (lanes 4 to 12). Binding complexes were resolved by electrophoresis in a 4% native polyacrylamide gel. For competition, 30 fold molar excess of the unlabeled DNA homologous to A probe was added to the incubation (lane 1). F, free probe; C8 and C9, gel retardation complexes. B. Methylation interference assay of the complexes C8 and C9. The G residues with methylation effects on protein-bindings are shown by relative positions to the E1A cap site. C. Comparison of the sequences at potential E1A-F factor binding sites. The nucleotide sequences within BS9, 10, 13, and 15 are shown compared to homologous sequences found within factor-binding sites characterized from other promoters : E4 (29) and polyoma enhncer element A(30). The G residues with methylation effects are shown by dots, where binding factors have been mapped to these sequences using methylation interference assay. This leads to a consensus E1A-F binding sequence as shown.

Two different E1A enhancer elements interact with a common factor with the binding sequence of XGGAYGT(X=A, C; Y=A, T).

The binding sites of BS10 and BS13 overlapped an upstream copy of the E1A enhancer element I and the distal E1A enhancer element, respectively (Fig. 2). Probe B that contains the

BS10 site generated three gel retardation complexes (C3, C4 and C5) by incubation with the 0.4M NaCl fraction of heparin-agarose, as previously shown in Fig. 4A. Methylation interference analysis indicated that complex 5 was the result of protein-bindings to the BS9 and BS10 sites (Fig. 4B). Three G residues at -299, -298, and -295 on the BS10 site and at -269 and -268 on the BS9 site were in close contacts with the protein on the noncoding strand. Probe A that contains the BS13 site yielded multiple gel retardation complexes with the 0.4M NaCl fraction (Fig. 5A, lane 8). Methylation interference analysis revealed that complex 8 was due to bindings to the BS13 and BS15 sites and that complex 9 was due to bindings to the BS15 site (Fig. 5B). Two G residues at -343 and -342 on the BS13 and -386 and -385 on the BS15 were in close contacts with the protein. One G residue at -387 on the BS15 may be contactpoints with the protein (Fig. 5B). By comparing the BS9, 15, 10 (enhancer element I) and 13 (distal enhancer element), a consensus sequence XGGAYGT(X=A, C; Y=A, T) can be derived (Fig. 5C). The two adjacent G residues were distinct contactpoints with the protein in all of the four sites. The results suggested that these sites including two different elements of E1A enhancer interacted with a common protein.

Protein binding to the hyphenated symmetric sequence at the E1A cap site

The cap site region of E1A gene is highly conserved in human adenoviruses and contains hyphenated symmetric sequences (8)(Fig. 2). The DNase I footprint assay revealed E1A cap site-binding activities in the 0.6M and 1.5M NaCl fractions of heparin-agarose (Fig. 1A). When E1A cap site-containing C probe was incubated with each of heparin-agarose fractions, a distinct gel retardation complex (C2) was generated with the 1.5M NaCl fraction but not with 0.6M NaCl fraction (Fig. 6A, lanes 13 and 10). Methylation interference assay indicated that complex C2 represented protein-binding to the BS1 site: G residues at -6, -4, -3, +7, +9 and +11 on the noncoding strand and -9, -2, -1, +2, +4, and +12 on the coding strand were in close contacts with the protein (Fig. 6B). These G residues were within the hyphenated symmetry region containing the E1A cap site (Fig. 2).

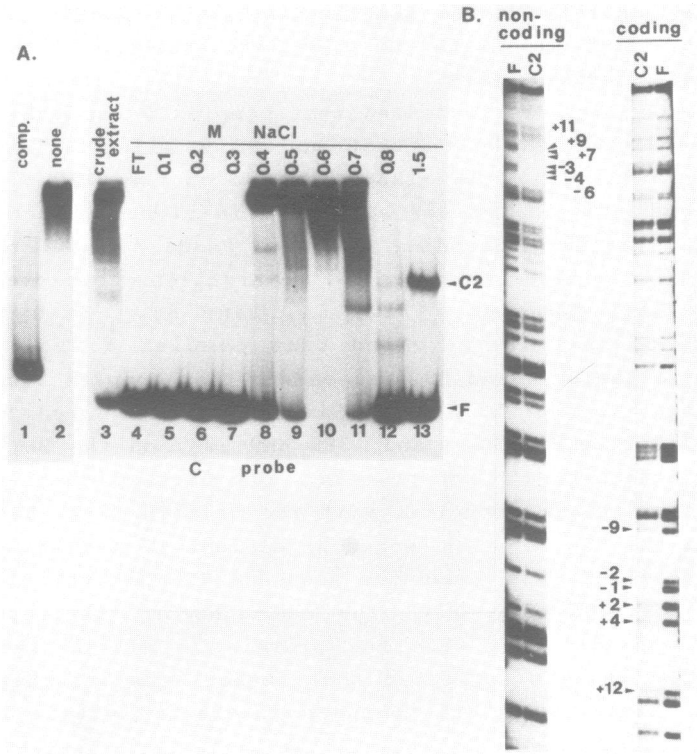


Figure 6. Protein-binding to the E1A cap site. A. Gel retardation assay of C probe containing the E1A cap site. Probe C (+50 to -157) was incubated with crude extracts and heparin-agarose fractions of HeLa cell nuclear proteins. Experimental procedures are described in Fig. 5A. B. Methylation interference assay of the complex 2 (C2). The G residues with methylation effects on protein-bindings are shown by relative positions to the E1A cap site.

DISCUSSION

The upstream region of the E1A gene contains cis-acting regulatory elements for transcriptions of the E1A gene and other early genes and the replication origin of viral DNA. We have identified twenty one binding sites of HeLa cell nuclear proteins on the upstream region of the Ad5 E1A gene. The binding sites and the proteins which interacted with them were summarized in Table I. Some of protein-binding sites overlapped the sites that have been shown to be important for efficient transcription of the E1A gene. The binding sites BS1-a and b were located at the hyphenated symmetry region

Table I. E1A upstream binding sites and putative factors

Binding site	Position of footprint	sequence	factor	M NaCl fraction of heparin-agarose	Regulatory element
BS1-a	-25 to +34			0.6M]	hyphenated symmetry
-b	-12 to +9			1.5M]	at the cap site(10)
BS2	-21 to -38			0.4M	TATA box(9,10)
BS3	-45 to -72			0.4M	TATA-like
BS4	-36 to -83			0.6M,0.7M	CCAAT box(10)
BS5	-91 to -99			0.6M,0.7M	
BS6	-165 to -178			0.7M	
BS7	-181 to -197	TGAAATCTGAATAA	E1A octa (NFIII, OBP10)	0.4M	partially overlap a downstream copy of enhancer element I(5,6)
BS8	-221 to -293			0.6M	enhancer element II(5,6) E2F-binding elements(7)
BS9	-257 to -273	CGGATGT	E1A-F	0.4M	
BS10	-289 to -304	AGGAAGT	E1A-F	0.4M	an upstream copy of enhancer element I(5,6)
BS11	-305 to -319			0.5M,0.6M	
BS12	-319 to -338	GTGACGT	ATF	0.6M,0.7M]	a distal enhancer
BS13	-334 to -346	CGGAAGT	E1A-F	0.4M]	element(4,12,33)
BS14	-363 to -384			0.5M,0.6M,0.7M	
BS15	-375 to -394	CGGAAGT	E1A-F	0.4M	
BS16	-391 to -414	GTGACGT	ATF	0.6M	(32,33)
BS17	-391 to -420	GGGCGG	SP1	0.7M	
BS18	-428 to -450	GTGACGT	ATF	0.6M	(32,33)
BS19	-452 to -464	CCAATATATGATAAT	NFIII (ORP-C)	0.4M	C domain of replication origin(19-21)
BS20]	-458 to -498	TGGATTGAAGCCA	NF1]	0.4M,0.5M	B domain(20-25)
BS21]			ORP-A]		A domain(21)

containing the E1A cap site. Viral mutants with deletions at the hyphenated symmetry yielded a lower level of E1A mRNAs three fold at early or late stages of infection (10). BS1 sites may play a role in E1A transcriptional regulation. BS2 coincided with the TATA box (8). Virus mutants with deletions of the TATA box decreased a level of E1A mRNAs moderately in vivo and shifted the cap site from a major site of E1A mRNAs to a set of minor sites found in wild type of E1A mRNAs (9, 10). The 5' end of one minor species of E1A mRNAs (9) was located 20 to 30 bp downstream from a AT strech (TATTATTATA, -59 to -50) which was within the BS3 (Fig. 2). The AT strech might be a potential TATA box in these TATA box-deletion mutants. BS4 overlapped BS3 and contained the CCAAT sequence GGTCAAAGT at -70 (8). Site BS7 was located at -190 and appeared to be dispensable for E1A transcription in Ad-infected cells (5, 6). We demonstrated that the -190 region must be a site of octamer motif-binding proteins, NFIII and OBP100 (E1A octa in Table I). It is likely in competition fashion that the -190 region also interacts with a minimal origin-binding factor. Site BS8 overlapped the enhancer element II which appeared to be important for transcription regulation of all early genes on the viral genome in infected cells (6). It also contained two binding sites of the E2F factor(7). Site BS10 and BS13 coincided with an upstream copy of the enhancer element I and the distal enhancer element, respectively. Comparison of the BS10, 9, 13 and 15 sites (this paper), the E4 promoter (29) and polyomavirus enhancer element A(5, 30) revealed a factor (E1A-F) with the consensus binding sequence XGGAYGT(X=A, C; Y=A, T). We suggested that two different E1A enhancer elements interacted with the E1A-F factor. This was in agreement with the result(12) that a potential binding protein to the distal E1A enhancer element also bound to the sequence around -299 (BS10 site in this study, Table I). The distal enhancer element also contained a potential site (BS12) for ATF factor, immediately downstream from the E1A-F site. The E1A-F and ATF factors are also juxtaposed around position -400 of the E1A promoter and positions -150 of the E4 promoter (29, 31), although their transcriptional significance remains unknown. Sites BS16 and 18 contained a sequence GTGACGT which has been shown to interact with the ATF factor (32, 33). Sites BS20, 19, and 21

were within the replication origin of Ad DNA, probably corresponding to the binding sites for the NF1 (20-25), NFIII (19-21), and ORP-A (21), respectively. Site BS21 overlapped the minimal origin (nucleotides 1 to 18)which is absolutely required for replication initiation of Ad DNA. Our competition binding assay suggested that the minimal origin, the proximal E1A promoter region, and the -190 region may interact with a common factor. We demonstrated that these three regions had a TATA box-like activity to position a site of transcription initiation approximately 30 bp downstream (manuscript in preparation). In early studies, one of the E1A upstream initiation sites of E1A transcripts were mapped around the -190 region in either cells infected with Ad2 (26) or transfected with plasmids containing Ad2 E1A gene (4). Recently, transcriptional activity in the reverse orientation was found at either termini of the Ad5 genome (35). Thus, it is possible that a minimal origin-binding factor is related to a TATA box factor. Biochemical characterization of these factors will be required to understand a relationship between minimal origin and TATA box factors.

ACKNOWLEDGEMENTS

We thank Masaki Iwabuchi and Koji Mikami for expert technical advice and Masaho Ishino and G. Chinnadurai for providing Ad5 E1A plasmids. This work was supported by Grants in Aid for Cancer Research and for Scientific Research from the Ministry of Education, Science and Culture, Japan.

REFERENCES

1. Nevins, J. R. (1981) Cell 26:213-220.
2. Berk, A. J. (1986) Annu. Rev. Genet. 20:45-79.
3. Van Ormondt, H., Maat, J., DeWaard, A. and Van der Eb, A. J. (1978) Gene 1:309-328.
4. Hen, R., Borrelli, E., Sassone-Corsi, P. and Chambon, P. (1983) Nucleic Acids Res. 24:8747-8760.
5. Hearing, P. and Shenk, T. (1983) Cell 33:695-703.
6. Hearing, P. and Shenk, T. (1986) Cell 45:229-236.
7. Kovesdei, I., Reichel, R. and Nevins, J. R. (1987) Proc. Natl. Acad. Sci. USA 84:2180-2184.
8. Van Ormondt, H. and Galibert, F. (1984) Current Topics in Microbiology and Immunology 110:73-142.
9. Osborne, T. F., Gaynor, R. B. and Berk, A. J. (1982) Cell 29:139-148.
10. Hearing, P. and Shenk, T. (1983) J. Mol. Biol. 167:809-822.

11. Dignam, J. D., Lebowitz, R. M. nad Roeder, R. G. (1983) Nucleic Acids Res. 11:1475-1489.
12. Barrett, P., Clark, L. and Hay, R. T. (1987) Nucleic Acids Res. 15:2719-2735.
13. Lichtsteiner, S., Wuarin, J. and Schibler, U. (1987) Cell 51:963-973.
14. Yoshida, K., Venkatesh, L. K., Mohan, K. and Chinnadurai, G. (1987) Genes Dev. 1:645-658.
15. Janson, L., Bark, C. and Pettersson, U. (1987) Nucleic Acids Res. 15:4997-5016.
16. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65:499-560.
17. Treisman, R. (1986) Cell 46:567-574.
18. Galas, D. J. and Schmits, A. (1978) Nucleic Acids Res. 5:3157-3170.
19. Pruijn, G. J. M., Van Driel, W., Van Miltenburg, R. T. and Van der Viliet, P. C. (1987) EMBO J. 6:3771-3778.
20. Pruijn, G. J. M., Van Driel, W. and Van der Vliet, P. C. (1986) Nature 322:656-659.
21. Rosenfeld, P. J., O'Neill, E. A., Wides, R. J. and Kelly, T. J. (1987) Mol. Cell. Biol. 7:875-886.
22. Nagata, K., Guggenheimer, R. A. and Hurwitz, J. (1983) Proc. Natl. Acad. Sci. USA 80:6177-6181.
23. Leegwater, P. A., Van Driel, W. and Van der Vliet, P. C. (1985) EMBO J. 4:1515-1521.
24. Rawlins, D. R., Rosenfeld, P. J., Wides, R. J., Challberg, M. D. and Kelly, T. J. (1984) Cell 37:309-319.
25. De Vires, E., Van Driel, W., Tromp, M., Van Boom, J. and Van der Vliet, P. C. (1985) Nucleic Acids Res. 13:4935-4952.
26. Garner, M. M. and Revzin, A. (1981) Nucleic Acids Res. 9:3047-3060.
27. De Vries, E., Van Driel, W., Van den Heuvel, S. J. L. and Van der Vliet, P. C. (1987) EMBO J. 16:161-168.
28. Sturm, R., Baumruker, T., Franza, B. R. Jr. and Herr, W. (1987) Genes Dev. 1:1147-1160.
29. Watanabe, H., Imai, T., Sharp, P. A. and Hamda, H. (1988) Mol. Cell. Biol. 8:1290-1300.
30. Yamaguchi, Y., Satake, M. and Ito, Y. (1989) J. Virol. 63:1040-1048.
31. Alieda, Leza, M. and Hearing, P. (1988) J. Virol. 62:3003-3013.
32. Hardy, S. and Shenk, T. (1988) Proc. Natl. Acad. Sci. USA 85:4171-4175.
33. Lin, Y. S. and Green, M. R. (1988) proc. Natl. Acad. Sci. USA 85:3396-3400.
34. Wides, R. J., Challberg, M. D., Rawlings, D. R. and Kelly, T. J. (1986) Mol. Cell. Biol. 7:864-874.
35. Ooyama, S., Imai, T., Hanaka, S. and Handa, H. (1989) EMBO J. 8:863-868.