Nucleic Acids Research, 1995, Vol. 23, No. 21 ***4457–4465***

DNA binding sites for the transcriptional activator/repressor YY1

Robin P. Hyde-DeRuyscher, Ezra Jennings and Thomas Shenk*

Howard Hughes Medical Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544-1014, USA

Received June 27, 1995; Revised and Accepted September 19, 1995

ABSTRACT

YY1 is a ubiquitously expressed zinc finger DNA binding protein. It can act as a transcriptional repressor or activator and, when binding at the initiator element, as a component of the basal transcription complex. Binding sites for YY1 have been reported in a wide variety of promoters and they exhibit substantial diversity in their sequence. To better understand how YY1 interacts with DNA and to be able to predict the presence of YY1 sites in a more comprehensive fashion, we have selected YY1 binding sites from a random pool of oligonucleotides. The sites display considerable heterogeneity, but contain a conserved 5′-CAT-3′ core flanked by variable regions, generating the consensus 5′-(C/g/a)(G/t)(C/t/a)CATN(T/a)(T/g/c)-3′, where the upper case letters represent the preferred base. This high degree of flexibility in DNA recognition can be predicted by modeling the interaction of the four YY1 zinc fingers with DNA and a detailed model for this interaction is presented and discussed.

INTRODUCTION

YY1 is a 414 amino acid zinc finger protein that is expressed in most, if not all, mammalian cell types. The human (1) and murine cDNAs (2–4) were cloned by groups studying the transcriptional regulation of different genes and the *Xenopus* counterpart has been cloned by virtue of its high degree of conservation with the mammalian counterparts (5). YY1 is a DNA binding transcription factor and it has been found to repress transcription from a variety of cellular promoters, including those from the immunoglobulin κ (4), skeletal α-actin (6–9), c-*fos* (10), ε-globin (11,12), α-globin (13), γ-interferon (14), GM-CSF (15), creatine kinase-M (16), *Pdha*-2 (17), α-1 acid glycoprotein (18), amyloid A1 (19), *Surf*1 and 2 (20) and β-casein genes (21,22). Several viruses have also been found to carry YY1 binding sites that have been shown to mediate transcriptional repression: Moloney murine leukemia virus (3), human papillomaviruses (23–25), Epstein Barr virus (26), human cytomegalovirus (27), human immunodeficiency virus (28), parvovirus (1,29) and adenovirus (30). Repression can also be modeled in artificial promoter constructs that contain YY1 binding sites (1,4).

The mechanism of repression in the case of the c-*fos* promoter has been proposed to involve YY1-mediated DNA bending, which could influence the ability of factors bound at upstream sites to interact with the basal transcription machinery (10). However, C-terminal segments of YY1 fused to a heterologous DNA binding domain can repress artificial promoter constructs that do not contain a YY1 binding site (1,4,31). Since the binding domain of the fusion protein does not inhibit transcription in the absence of the YY1 segment, YY1 must contain a repression activity within its C-terminal domain that does not depend on its ability to bend DNA. This domain might encode an intrinsic repression activity or it could interact with another protein that contains such an intrinsic activity.

In several cases the site at which YY1 binds and represses transcription overlaps the recognition site for a second inducible factor that activates transcription (6,7,10,19,32,33). Apparently, YY1 occupies its binding site and represses transcription until the activating factor is induced and successfully competes with YY1 for occupancy of the site. This displacement strategy results in a strong induction of transcription, since relief of YY1-mediated repression and transcriptional activation would occur simultaneously.

In some promoters YY1 binding sites can positively modulate transcription when located near the site of transcriptional initiation (2,8,18,33–38). Presumably the context in which YY1 binds influences its function, but it may also behave differently when bound near the start site of transcription. In fact, YY1 can activate transcription at several initiator elements (39,40). Under the appropriate *in vitro* conditions YY1, TFIIB and RNA polymerase II can mediate transcription from the initiator of the adeno-associated virus P5 promoter in the absence of the other known auxiliary factors, including TFIID (41).

The activity of YY1 can be modulated by the adenovirus E1A oncoprotein (1), which binds to YY1, relieving YY1-mediated repression (1,31). Like E1A, the c-myc protein (42) and the B23 protein (43) bind to YY1 and abrogate its ability to repress transcription. YY1 has also been shown to interact with p300 (44) and Sp1 (45,46).

YY1 contains four C_2H_2-type zinc fingers and structural analysis of other zinc finger proteins has demonstrated that each zinc finger interacting with DNA contacts 3–5 bp (47,48). Assuming that all four YY1 zinc fingers contact DNA, YY1 should recognize a binding site comprised of at least 12 bp and DNase I footprinting, as well as methylation interference analyses

* To whom correspondence should be addressed

(1,2,4,6,10,11,13,17,35,39,49), are consistent with a protein–DNA interaction that spans at least 12 bp. However, comparison of known YY1 binding sites indicates that only 3 bp of the recognition site are invariant, making it very difficult to scan transcriptional control regions and predict with any confidence the presence of YY1 binding sites. Therefore, we decided to identify the range of DNA sequences to which YY1 can bind, so that it would be possible to identify functional YY1 binding sites by searching DNA sequences for members of a substantial set of known binding sites, rather than an ambiguous consensus site.

Starting from a pool of oligonucleotides containing random sequences, we utilized a glutathione S-transferase–YY1 (GST–YY1) fusion protein to affinity purify oligonucleotides with YY1 binding sites. After six sequential rounds of selection, followed by amplification of bound oligonucleotides by PCR, sequence analysis of the binding sites revealed a core 5′-CCAT-3′ sequence surrounded by variable sequences. All but one of the selected sequences tested interacted with YY1 in a band shift assay and mediated repression. YY1 repressed model promoters containing these sites, irrespective of their orientation. Comparison of the YY1 binding sequences with the binding sites for zinc finger proteins for which a protein–DNA co-crystal structure has been solved indicated that YY1 most likely interacts with a 12 bp sequence. A computer search of a promoter database revealed the presence of YY1 binding sites in a wide variety of viral and cellular promoters, many of which overlap with sites for other known transcription factors.

MATERIALS AND METHODS

Selection of YY1 binding sites

The YY1 coding region was cloned into pGEX2T (Pharmacia) to produce a GST–YY1 fusion protein. After induction with IPTG a lysate was prepared from *Escherichia coli* DH5α cells containing either pGEX2T or pGEX2T-YY1 by sonication in NETN buffer (100 mM NaCl, 1 mM EDTA, 20 mM Tris–HCl, pH 8.0, 0.5% NP-40) and 100 µl aliquots were frozen in liquid N_2. Glutathione–Sepharose (Pharmacia) was washed three times with NETN buffer and used to isolate fusion protein by incubating 100 µl *E.coli* lysate with 50 µl washed beads at 4°C for 30 min while rotating. The beads were pelleted by centrifugation and washed twice with 1 ml NETN buffer and twice with 1 ml 1× binding buffer (12 mM HEPES, pH 7.9, 60 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.5 mM EDTA, 0.05% NP-40, 50 µg/ml bovine serum albumin, 10% glycerol). The final pellet was resuspended in 100 µl 1× binding buffer.

A 55 nt DNA was synthesized that contained a central 15 nt random sequence flanked by sequences for the binding of PCR primers: 5′-CTGTCGGAATTCGCTGACGT$(N)_{15}$CGTCTTATCGGATCCTACGT-3′. Two primers were synthesized with the sequences 5′-CTGTCGGAATTCGCTGACG-3′ (upstream primer) and 5′-ACGTAGGATCCGATAAGACG-3′ (downstream primer). Double-stranded oligonucleotide was generated by an initial PCR reaction containing 10 ng 55mer, 0.1 µg each primer, 1× PCR buffer (10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 0.2 mM of each deoxynucleoside triphosphate (Pharmacia) and 5 U Amplitaq DNA polymerase (Perkin-Elmer Cetus). Amplification (94°C, 30 s; 65°C, 1 min; 72°C, 1 min) was carried out for 25 cycles in a volume of 50 µl and then the first round of capture was initiated using the amplified 55mer without further purification.

For the first round of capture 1 µl amplified PCR mixture was mixed with an excess of GST–YY1 bound to glutathione–Sepharose beads in 100 µl 1× binding buffer containing 1 µg poly(dI-dC:dI-dC) (Pharmacia), which was included in all of the rounds of capture as a non-specific competitor. This mixture was incubated at room temperature for 30 min with continual rotation. The GST–YY1–DNA complexes on beads were collected by centrifugation at 6000 r.p.m. for 10 s and the pellet resuspended, washed three times with 1 ml 1× binding buffer and finally resuspended in 50 µl 1× binding buffer. The bound oligonucleotides were extracted from the GST–YY1 on beads by addition of 148 µl TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA) and 2 µl 10% SDS with incubation of the mixture at 95°C for 10 min. The mixture was extracted with 200 µl phenol:chloroform (1:1) and the DNA precipitated using 20 µg glycogen as carrier. The DNA was resuspended in 20 µl H_2O and 1 µl of this was used for amplification using the conditions given above, except that only 10 cycles were used (a greater number of cycles resulted in the production of multimers of the original 55mer). Six rounds of capture and amplification were performed. After the last amplification 10 µl amplified DNA were subjected to an additional 25 cycles of amplification to facilitate cloning of the selected binding sites.

Cloning and analysis of YY1 binding sites

The final product of the capture and amplification selection procedure was purified from a 4% NuSieve agarose gel (FMC) using a Spin-X column (Costar). The purified DNA was digested with *Eco*RI and *Bam*HI, gel purified and cloned into *Eco*RI/*Bam*HI-digested pBluescript (Stratagene). After transformation of *E.coli*, white colonies were picked, 1 ml DNA preparations were made using Magic Mini-prep resin (Promega) and cloned oligonucleotides were sequenced. Clones that were chosen for transient gene expression analysis were subcloned from pBluescript into pTILUC by PCR amplification of the insert using two primers with the sequences 5′-CTGGCGGAATTCGCAGTCGT-3′ and 5′-ACGTAGGATCCGATAAGACG-3′. The first primer was altered so that the product would no longer contain a potential match to an ATF binding site which we identified in the original primer sequence and which could have complicated subsequent analysis. The PCR products were then treated with T4 DNA polymerase, digested with *Bam*HI, purified on a 4% NuSieve agarose gel and cloned into *Bgl*II/*Sma*I-digested pTILUC. All of the final clones were verified by DNA sequencing.

Band shift analysis was done as previously described (1). Briefly, reactions were in a final volume of 14 µl 1× binding buffer containing 1 µg poly(dI-dC:dI-dC) (Pharmacia), 10 fmol ^{32}P-labeled probe DNA and specified amounts of competitor oligonucleotides. Complexes were separated on 4% acrylamide (29:1 acrylamide:bis-acrylamide) in 0.5× Tris–glycine buffer. Probes for individual selected binding sites were prepared using ^{32}P-labeled primers (described above for cloning into pTILUC) to amplify the binding sites from 10 ng plasmid DNA in a volume of 25 µl. Signals were quantified using a Molecular Dynamics PhosphorImager.

HeLa and PYS-2 cells were maintained in DMEM supplemented with 10% calf serum. Each well of a 6-well culture plate was seeded with 3×10^5 cells the night before transfection, the medium was changed the next morning and transfections were carried out the same afternoon by the calcium phosphate precipitation method using DNA (10 µg reporter plasmid) purified on Promega Magic Resin. The medium was changed

Table 1. Sequences of oligonucleotides captured by YY1 CASTing

Sequence	Clone	Orientation
acgtAAACGCGCCATTTTGcgtcttatcggat	19	-
gacgtATTGCGCCATTTTGTcgtcttatcgga	10	-
ccgataagacgCCATTTTAAGTCCTacgtca	11	+
gataagacgCGCCATTTTGTGTTacgtcagcg	55	+
taagacgTCCGCCATTTTGTGTacgtcagcga	88	+
ccgataagacgCCATTTTGAAATACCacgtca	29	+
ccgataagacgCCATTTTGAATGCACacgtca	63	+
gacgtACGTCGCCATTTTGAcgtcttatcgga	75	-
ccgataagacgCCATTTTGGAGCATTacgtca	58	+
ccgataagacgCCATTTTACGGATGGacgtca	92	+
ccgataagacgCCATTTTACTCATGTacgtca	33	+
ccgataagacgCCATTTTAAAACGCTacgtca	25	+
ttcgctgacgtCCATTTTTAACATGTcgtctt	50	-
ttcgctgacgtCCATTTTTGTCATGTcgtctt	28	-
ttcgctgacgtCCATTTTAGTTATGGcgtctt	34	-
ataagacgCGTCCATTTTGTTGTacgtcagcg	72	+
ttcgctgacgtCCATTTTGTTCCTCCcgtctt	61	-
cgctgacgtAGCCATTTTCTTTCAcgtcttat	37	-
ataagacgTCGCCATCTTGTCTTacgtcagcg	20	+
taagacgGTCGCCATCTTGTCCacgtcagcga	77	+
ccgataagacgCCATCTTGTATTTTGacgtca	25	+
ccgataagacgCCATCTTGATGTCTCacgtca	62	+
ccgataagacgCCATCTTGATACATCacgtca	74	+
ccgataagacgCCATCTTGCCTACTacgtcag	70	+
gataagacgCGCCATCTTTGTCTacgtcagcg	81	+
ccgataagacgCCATCTTTTAACGCAacgtca	42	+
ttcgctgacgtCCATCTTTAATATGTcgtctt	31	-
ataagacgCATCCATCTTGACTTacgtcagcg	91	+
gctgacgtCGGCCATCTTGTCTGcgtcttatc	80	-
acgtGTAAGCGCCATGTTGcgtcttatcggat	83	-
acgtATGTCCGCCATGTTGcgtcttatcggat	44	-
gctgacgtACGCCATGTTGcgtcttatcggat	68	-
ccgataagacgCCATGTTGGGACCTAacgtca	65	+
attcgctgacgCCATGTTGGCTGAGcgtctta	60-,64	-
ccgataagacgCCATGTTAACTAATCacgtca	45	+
ttcgctgacgtCCATGTTGAGTTTCcgtctta	16	-
ttcgctgacgtCCATGTTAGCAATGGcgtctt	78	-
cgataagacggCCATGTTGTCTTGAacgtcag	26	+
ccgataagacgCCATATTCCTCCATTacgtc	15	+
ccgataagacgCCATATTGTCTATAacgtcag	43	+
gacgGGAGCCGCCATATTTacgtcagcgaatt	46	+
ttcgctgacgtCCATATTGTAATGGcgtctt	79	-
ccgataagacgCCATATGCAATTTCCacgtca	36	+
ccgataagacgCCATTATTGTACTGTacgtca	56	+
ccgataagacgCCATTATCATCATGGacgtca	24	+
ccgataagacgCCATTACGGAATACAacgtca	18	+
acgtAAATCCGCCATTTGCcgtcttatcggat	32	-
ttcgctgacgtCCATCTTGATAATGGcgtctt	31	-/+
ttcgctgacgtCCATATTAAACATGGcgtctt	39	-/+
ttcgctgacgtCCATATTGCAAATGGcgtctt	64	-/+
ccgataagacgCCATTTGTAATATGGacgtca	49	-/+
ccgataagacgCCATTGCAATCATGGacgtca	23	+/-
ttcgctgacgtCCATGATGTAAAATGTcgtcc	48	-/+
ccgataagacgCCATTTTCATCATGGacgtca	40	+/-
ccgataagacgCCATATTTTCAATGGacgtca	85	+/-
cgataagacgCGAAGTATTAACTAAacgtcag	14	+
taCGCCATtTTg gtgt cagt Aga g ca cc a c	Composite	

Lower case letters indicate sequences in the constant primer regions and upper case letters indicate the random region of the oligonucleotide. The 9 bp conserved sequence is separated from the flanking regions. The composite site shown at the bottom shows all of the bases found at each position in clones where single binding sites are contained entirely within the random sequence region of the oligonucleotide.

18 h later and the cells harvested 40 h after the DNA was added. Extracts for luciferase assays were made using 150 µl/well Promega Reporter Lysis buffer. Aliquots of 100 µl extract were used for each assay using 100 µl each of luciferase assay buffer (reagent A) and enhanced luciferase substrate (reagent B) from Analytical Luminescence Laboratory. Assays were performed using a 10 s measurement on a Analytical Luminescence Laboratory Monolight 2010 luminometer.

Computer searches

Eukaryotic Promoter Database, release 32 (50) was downloaded to a Macintosh computer. The FASTA file was converted into Filemaker Pro and searched for the binding sites indicated in Table 1.

RESULTS

Selection of YY1 binding sites from a pool of random oligonucleotides

To confirm that the GST–YY1 fusion protein bound to glutathione–Sepharose beads would interact with its DNA binding site in a specific fashion we mixed GST–YY1 protein immobilized on beads with ^{32}P-labeled oligonucleotides that contained known binding sites for several different factors. The beads were then collected by centrifugation, washed several times and the retained ^{32}P-labeled DNA was quantified. Only oligonucleotides that contained a YY1 binding site were captured by the fusion protein (data not shown).

To isolate YY1 binding sites from a pool of random oligonucleotides the GST–YY1 fusion protein bound to beads was mixed with a pool of oligonucleotides containing a core of 15 random bases flanked by two primer binding sites. The GST–YY1–DNA complexes on beads were collected by centrifugation, washed in binding buffer and the bound DNA eluted from the beads and amplified by PCR. This amplified pool was again bound to the GST–YY1 beads and the entire process repeated six times (Fig. 1). As a preliminary control for specificity we also used GST alone bound to beads and the amplified product from each cycle was analyzed by electrophoresis on an agarose gel. Only the beads containing the GST–YY1 fusion protein captured DNA (data not shown). We found that amplification of the DNA for >10 rounds produced multimers of the 55mer PCR product, so we limited the amplification to 10 rounds after each capture cycle. The final round of amplified material was then cloned and 56 individual clones were sequenced (Table 1). The captured binding sequences fell into three broad classes: those with single binding sites (47 clones), those that appeared to have two binding sites in opposite orientations (eight clones) and those that did not have a recognizable binding site (one clone).

All of the single binding sites contained a conserved core of 5′-CCAT-3′ flanked by upstream and downstream regions exhibiting a considerable degree of sequence flexibility. The 2 bp immediately 5′ of the conserved core are somewhat flexible, but 5′-CG-3′ is favored (34 out of 47 clones, 72%). Analysis of these two 5′ base pairs is complicated by the non-random primer flanking region, which ends with a 5′-CG-3′ in one orientation. This appears to have selected for binding sites located toward the edge of the random sequence stretch. In addition, the non-random primer flanking sequences in the opposite orientation end with GT, the second most common sequence present at this position (nine out of 47 clones, 18%). In order to avoid biasing the results, the frequencies indicated for individual base pairs at these two positions in Table 1 were calculated only for those binding sites in which they occurred within the 15 bp random stretch. Nevertheless, even when the analysis is restricted to random sequences, 5′-CG-3′ was found immediately 5′ of the 5′-CCAT-3′ core in 13 of 17 clones (76%). The base pair immediately 3′ of the 5′-CCAT-3′ core is the most variable position in the binding site, but is most often a T (22 out of 47 clones, 47%). The next 2 bp on the 3′ side are somewhat flexible, but are predominantly 5′-TT-3′ (42 out of 47 clones, 89%). This core site of 5′-CGCCATTTT-3′ is often preceded on the 5′ side by a C (seven out of 14, 50%) and followed on the 3′ side by a G (28 out of 47 clones, 60%).

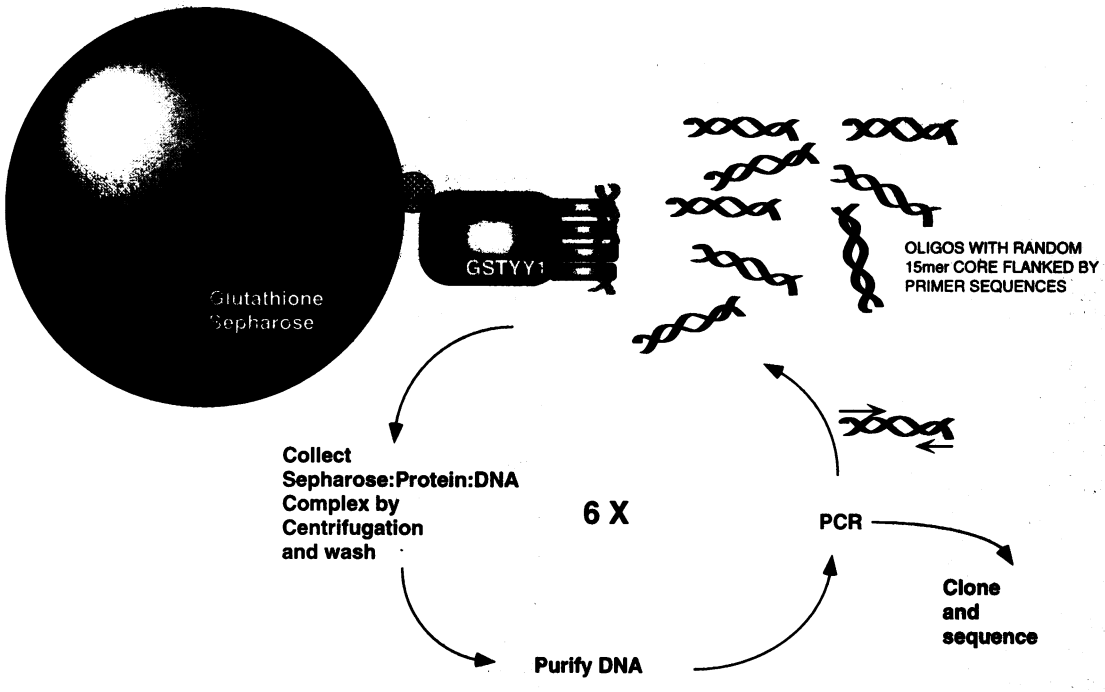

Figure 1. Flow chart depicting the procedure used to isolate YY1 binding sites.

All but one of the oligonucleotides with two binding sites contain two conserved 5′-CCAT-3′ motifs positioned in opposite directions. The exception contains only one 5′-CCAT-3′ sequence with 5′-ACAT-3′ in the opposite orientation, a sequence found at a known YY1 site in the adeno-associated virus P5 promoter (1). Without synthesizing oligonucleotides for each inverted binding site to separate them it is impossible to know if only one or both of these sequences bind YY1. Therefore these sequences were not used to screen databases or to develop a consensus binding site. Clone 14 (Table 1) lacked a 5′-CCAT-3′ or 5′-ACAT-3′ core and did not contain any clear homology to other known YY1 binding sites.

Analysis of selected sites

To demonstrate that YY1 can, indeed, bind to the selected sites a representative group was chosen for DNA band shift analysis. Oligonucleotide probes were prepared by PCR amplification of the cloned binding sites from plasmid DNA using ^{32}P-labeled primers. The resulting probes were incubated with either His–YY1 or GST–YY1 fusion proteins prepared from *E.coli* or with a HeLa nuclear extract containing native YY1. Each of the sources of YY1 provided the same results and a representative band shift assay using His–YY1 fusion protein is displayed in Figure 2A. All of the selected clones bound to His–YY1 except clone 14, the clone that did not contain a clear homology to the other binding sites. This clone bound less His–YY1 than a randomly chosen clone that had not gone through the selection procedure. Thus it appears that clone 14 is a contaminant that has been carried through the six cycles of sequential selection and amplification.

The relative efficiency with which YY1 bound to each site was estimated by comparing the amount of shifted complex formed with different oligonucleotide probes in Figure 2A. All of the probes were made using the same ^{32}P-labeled primers and thus have the same specific activity, making comparison of the shifted complexes produced in the presence of a constant amount of His–YY1 protein an easy way to assess the relative efficiency of binding. We observed a 6-fold difference in the efficiency of binding to His–YY1 for the sites tested (Fig. 2B) and similar results were obtained using GST–YY1 fusion protein or native YY1 present in a HeLa cell extract (data not shown). The relative binding efficiencies could be predicted by the number of individual clones isolated which contained that binding site. The site binding with the highest efficiency was isolated 12 times, while the site binding with the lowest efficiency was isolated only once (Table 1).

To determine if the selected sites can affect transcription, some of the sites were cloned upstream of a minimal promoter controlling luciferase gene expression. The minimal promoter was comprised of the TATA motif from the adenovirus major late promoter and the initiator element from the terminal deoxynucleotidyl transferase gene. Given the different locations of the YY1 recognition site in the selected oligonucleotides, the resulting reporter plasmids contained the YY1 binding site at somewhat different distances (–70 to –80) and orientations relative to the start of transcription. The reporter plasmids were transfected into HeLa cells and the levels of luciferase that they produced were compared with the amount produced by the parent vector lacking a YY1 binding site (Fig. 3A). Each of the sites with which YY1 interacted in the band shift assay repressed expression from the reporter plasmid; none of the inserted YY1 binding sites activated expression. Repression ranged from a factor of 5 to 100 and occurred with the YY1 binding site inserted in either orientation. There was no correlation between the efficiency of binding measured by band shift and the degree of repression. The DNA sequence present in clone 14, which did not score in the YY1 band shift assay, did not repress transcription, confirming our conclusion that it is a random contaminant which survived the selection procedure.

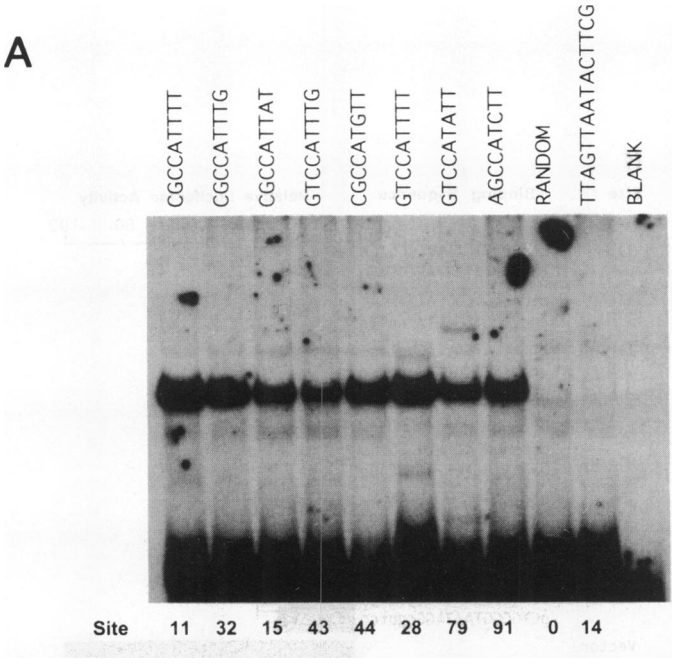

Figure 2. YY1 binding to representative recognition sites. (**A**) Electromobility shift analysis. Oligonucleotide probes (55 bp) for each clone were prepared by PCR amplification of cloned recognition sites using ^{32}P-labeled primers. Probes were mixed with His–YY1 fusion protein and binding was assayed by electrophoresis. The binding site number is given with reference to Table 1 and the nine base core sequence of each probe is shown above each lane. The lane labeled random received a reaction containing a mixture of oligonucleotides isolated by amplification from the starting material for the CAST. The lane labeled blank received a reaction containing no probe, generated by a PCR reaction that did not receive template DNA. (**B**) The radioactivity in YY1-specific shifted complexes was quantified using a PhosphorImager and plotted relative to that for site 11.

The screen for repression mediated by YY1 binding sites indicated that sites in both possible orientations relative to the start site sponsored repression. To more rigorously test the orientation dependence of repression we placed two of the sites upstream of the SV40 early promoter in both orientations at identical distances from the transcription start site. We also prepared constructs in which the YY1 binding site was moved half a helical turn from its original location, to test the possibility that the activity of YY1 would be altered as it was moved from one face of the helix to the other. In each case the sites repressed transcription between 4- and 7-fold, indicating that repression is orientation independent and insensitive to changes in the helical face occupied by YY1 relative to the start site (Fig. 3B). This experiment also demonstrated that the YY1 binding sites can repress in the context of a second promoter, the SV40 early promoter.

To determine if the repression observed in HeLa cells was cell type specific, several of the constructs were transfected into PYS-2 cells. These cells have previously been reported to activate transcription through YY1 (35). In the minimal promoter constructs used here all of the sites repressed transcription from 2- to 10-fold (Fig. 4). In contrast, a Gal4–YY1 fusion protein activated transcription of a reporter plasmid containing four Gal4 binding sites upstream of the thymidine kinase promoter in PYS-2 cells. The luciferase assay generated 29 226 light units in the presence of Gal4 alone, versus 193 288 light units in the presence of Gal4–YY1, a 6.6-fold activation. This has been previously reported for the fusion protein in PYS-2 cells (35) and is opposite to the results found in HeLa cells with the fusion protein (31) or in PYS-2 cells (Fig. 4) assaying endogeneous YY1 function at a YY1 binding site upstream of a minimal promoter. Clearly, the fusion protein can function differently from the native YY1 protein.

Prevalence of YY1 binding sites

To evaluate the prevalence of YY1 binding sites we first searched a portion of the Genbank-EMBL database for sites in mammalian DNA using the set of YY1 sites shown in Table 2. Each of these sites has been shown to be a physiologically active YY1 binding site in at least one promoter context. This search identified 5954 binding sites in 56 000 000 bp of sequence searched. To focus the analysis we searched the Eukaryotic Promoter Data Base maintained at the Institut Swisse de Recherches Experimentales sur le Cancer by Philipp Bucher. This database contains 778 entries from vertebrate and viral genes with sequences from –500 to +100 bp of the transcription start site for RNA polymerase II-transcribed genes. The search found 46 sites in the promoters of 624 vertebrate genes and 37 sites in the promoters of 154 viral genes. If these sites occured at random in our searches we would expect to find 51 hits, rather than the 83 YY1 binding sites observed. All 83 hits correspond to known YY1 binding sites, so we can conclude that YY1 binds to the control regions of a wide variety of genes. The search did not reveal any particular class of genes with a predilection for YY1 binding sites.

DISCUSSION

We used a GST–YY1 fusion protein to isolate YY1 binding sites from a pool of random oligonucleotides. A 5′-SKCCATNTT-3′ consensus sequence was deduced, with 5′-CGCCATTTT-3′ being the site captured most frequently (Table 1). These results are in agreement with previous compilations of YY1 sites (7,51). All of the selected sites that were demonstrated to bind YY1 by band shift assay contained a conserved 5′-CCAT-3′ core (Fig. 2). However, the invariant core must be reduced to 5′-CAT-3′ if one considers the 5′-ACAT-3′ core sequence in the YY1 binding site centered at –60 in the adeno-associated virus P5 promoter and the 5′-TCAT-3′ core in the ε-globin promoter (Table 2). The sequences located on the 5′ and 3′ sides of the core were relatively flexible. This variability in binding sites potentially allows YY1 to bind and influence transcription within a wide variety of promoters. This flexibility in binding motifs might also enable YY1 to compete for binding with many transcription factors at

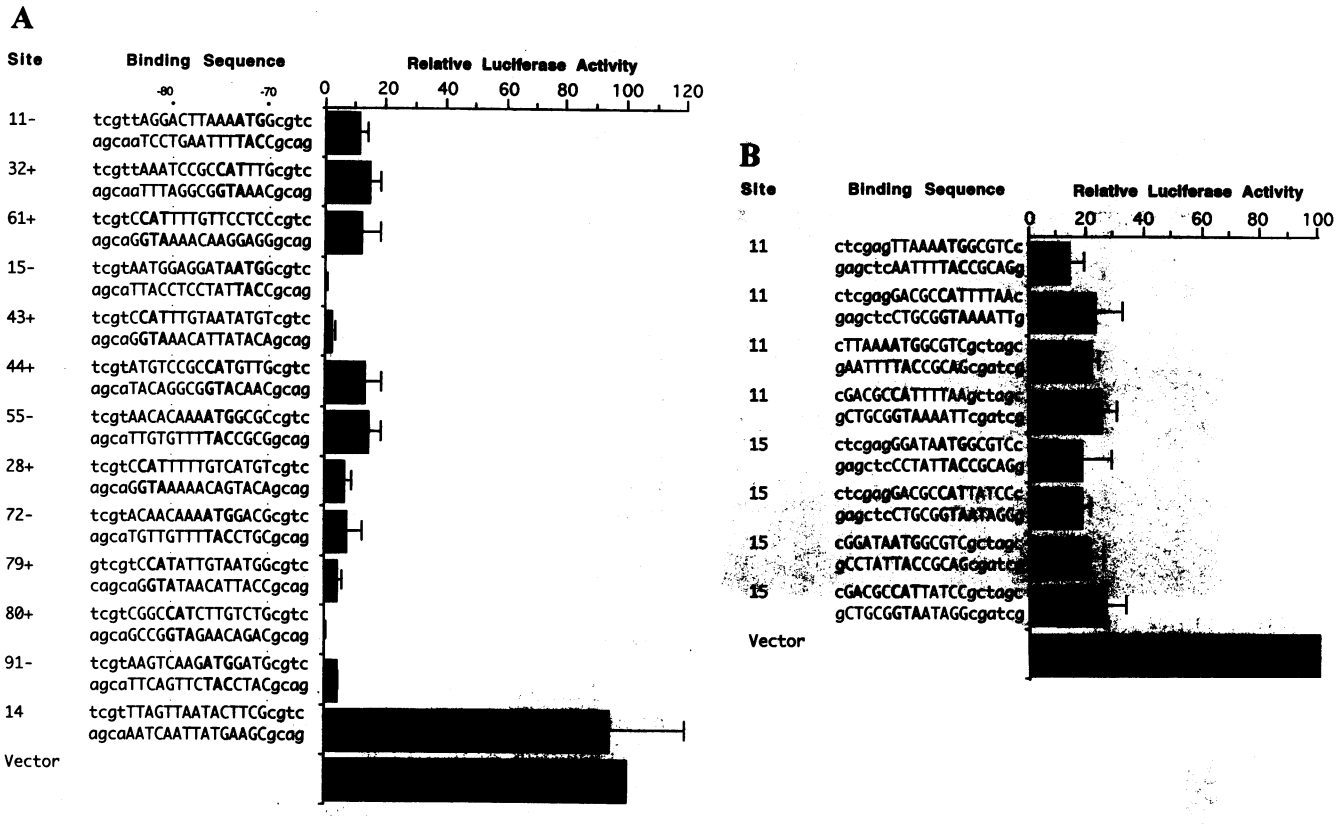

Figure 3. Luciferase activity from reporter plasmids containing various YY1 binding sites in HeLa cells. The indicated binding sites were cloned into basal promoters, transfected into HeLa cells and luciferase activity expressed by the clones with YY1 binding sites was compared with that produced by the vector without any binding site. (**A**) Luciferase activity from pTiLUC containing various binding sites. The central CAT core for each site is presented in bold. The numbers above the sequence represents the distance from the start site of transcription. The activities are the average of four independent experiments and error bars are shown for each construct. (**B**) Luciferase activity from the pGL2 promoter containing synthetic oligonucleotides representing YY1 binding sites 11 and 15. The binding sites were cloned to allow direct comparison of the two possible orientations and the effect of displacing sites half a helical turn. The activities are the average of three independent experiments and error bars are shown for each construct.

overlapping recognition sites. This competition could result in more stringent transcriptional regulation than would otherwise be possible (discussed below).

We used a GST–YY1 fusion protein bound to glutathione–Sepharose beads as the matrix for the CASTing experiment (Fig. 1). In previous studies gel shift assays (52,53), immunoprecipitation (54) and affinity chromotography (55) have been used to separate DNA–protein complexes from unbound DNA. These methods generally have allowed a significant fraction of contaminating DNA lacking specific binding sites to be carried through multiple rounds of selection. In this study only one out of 63 clones analyzed contained a site which did not bind to YY1, suggesting that the use of fusion proteins on beads is a highly efficient method to isolate binding sites of interest free from contaminating sequences. It is also possible, however, that we have carried out the selection under overly stringent conditions or for more cycles than optimal. As a result, we might have selected against classes of recognition sites to which YY1 binds with lower affinity than to the sites we selected.

A search of the Eukaryotic Promoter Database revealed that there are YY1 sites in the putative transcriptional control regions of a wide variety of genes; 83 YY1 recognition sites were found in a search of 778 promoters. There does not appear to be one class of genes represented, i.e. genes of the immune system, housekeeping genes, TATA-less genes, etc., in greater abundance than others. This is not unexpected, since YY1 appears to be ubiquitously expressed and probably does not by itself determine the specificity of expression of any one class of gene.

Mechanism of YY1 action

Natesan and Gilman found that the activity of the YY1 site in the c-*fos* promoter is orientation dependent, that YY1 binding to the c-*fos* promoter can bend DNA and that YY1 did not repress transcription from the c-*fos* promoter in the absence of upstream enhancer elements (10). These observations led them to propose that YY1 modulates c-*fos* transcription by bending DNA to modulate contacts between other proteins that interact within the promoter and enhancer domains. In contrast, our study has found that YY1 can repress transcription regardless of orientation from a synthetic basic promoter construct containing only a TATA box and initiator element and from the SV40 early promoter (Fig. 3). Apparently, the mechanism of repression by YY1 bound to the c-*fos* promoter is at least in part different from that by which YY1 inhibits expression from the promoters we have tested. If this is true, then the difference must result from the context in which YY1 binds. It is unlikely that the sequence of the YY1 binding site influences repression activity, since the binding sites used in this study were selected only for their ability to bind YY1 and all of the sites tested mediated orientation-independent repression. This

observation, together with earlier work showing that a C-terminal segment of YY1 fused to a Gal4 DNA binding domain can repress transcription from promoters lacking a YY1 binding site (1,4,31), argues that the DNA bending model does not account for many YY1-mediated repression events. We favor a model for repression that is not dependent on DNA bending, in which YY1 either contains an intrinsic repression domain or a domain that binds to other proteins with intrinsic repression activity. Perhaps the YY1 repression domain itself, or a protein with which it interacts, competes for and blocks a critical interaction that must occur between constituents of the basal transcription complex. This seems a plausible hypothesis, given the ability of YY1 to participate in the transcriptional initiation reaction when bound at an initiator element (34,39–41).

Table 2. Derivation of a nine base consensus core sequence for all of the known YY1 binding sites

	Site	Source
1	CGCCATTTT	This study, M-MLV LTR (3), hCMV IE enhancer (30)
2	GTCCATTTT	This study
3	AGCCATTTT	This study
4	CGCCATCTT	This study, B19 P6 (32), IAP (22)
5	GTCCATCTT	This study
6	AGCCATCTT	This study, *Pdha-2* (18), *Surf1* (21), *Surf2* (21)
7	GGCCATCTT	This study, IgH enhancer (4), rpL30 (2), HSV1 VP5 (28), LINE-1 (16)
8	CGCCATGTT	This study
9	GTCCATGTT	This study
10	GGCCATGTT	This study
11	CGCCATATT	This study, α-actin (6–9)
12	CGCCATATG	This study
13	CGCCATTAT	This study
14	CGCCATTAC	This study
15	CGCCATTTG	This study
16	CGACATTTT	AAV P5(–60) (1)
17	CTCCATTTT	AAV P5(+1) (1)
18	CTCCATCTT	Igκ3′ enhancer (4)
19	TGCCATCTG	*rpL32* (2)
20	GGCCATCCG	*rpL32* (2)
21	TGACATATT	δ-globin (11,12)
22	TATCATTTT	ε-globin (11,12)
23	TCCCATTCT	ε-globin (11,12)
24	CTTCATCAT	ε-globin (11,12)
25	AGCCATATG	EBV BZLF1 (28)
26	GTCCATATT	c-*fos* (10)
27	GACCATTTT	c-*myc* (8)
28	CGCCATGTA	c-*myc* (8)
29	GCCCATCTT	*Cox*Vb (39)
30	CGCCATACT	α-globin (13)
31	AACCATTTT	β-casein (23,24)
32	TTTCATTAA	HPV-18 (25)
33	GTTCATTTG	HPV-16 (26)
34	GTTCATTTG	HPV-16 (26)
35	ACCCATGTG	HPV-16 (26)
36	CACCATTTT	Ad12 (34)
37	CCCCATACA	Creatine kinase-M (17)
38	TGCCATTCT	Interferon-γ (14)
40	CACCATGTC	Serum amyloid A1 (20)
41	GGCCATTTA	hCMV IE enhancer (31)

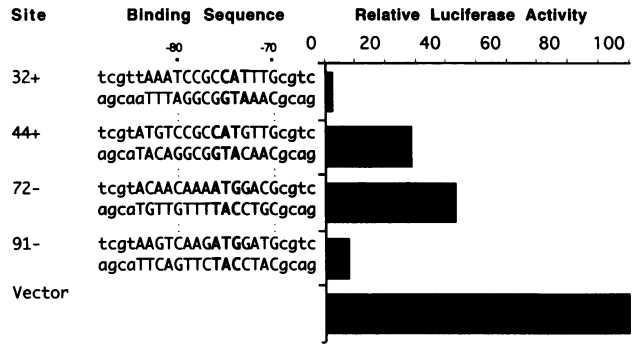

Figure 4. Luciferase activity from reporter plasmids containing various YY1 binding sites in PYS-2 cells. The indicated binding sites are a subset of the YY1 sites assayed in HeLa cells in Figure 3A. Reporter constructs were transfected into PYS-2 cells and luciferase activity expressed by the clones with YY1 binding sites was compared with that produced by the vector without any binding site.

In some cases YY1 might repress in part by competing for DNA occupancy with a positive acting transcription factor with an overlapping binding site. Even though all of the YY1 binding sites share a 5′-CAT-3′ core, the base pairs on either side of the conserved core are quite variable, offering YY1 sites the ability to overlap with a wide range of other DNA binding sites. Indeed, YY1 and the serum response factor (SRF) compete for overlapping sites within the c-*fos* promoter (6,7,9). This competition results in an antagonistic effect of the two proteins. In the absence of SRF, YY1 can bind to the promoter and repress transcription, and when SRF is induced, it can compete with YY1 for access to the DNA and activate transcription if it wins the competition. YY1 exhibits very rapid on and off rates when binding of purified protein is assayed *in vitro* (<10 s; Hyde-DeRuyscher and Shenk, unpublished). Since the concentration of YY1 within cells does not appear substantially higher than for many other transcription factors, it seems likely that many transcription factors exhibiting slower off rates could successfully displace YY1 from an overlapping binding site. This competition between positive and negative factors could result in a net repression or activation depending on the relative concentration of the factors and the competition could provide for a more stringent control of gene expression than could be obtained if induction involved the binding of an activator without the simultaneous displacement of a repressor protein.

YY1 can activate transcription from several promoters when it binds at the transcriptional initiator element (34,39–41). In a reconstituted, cell-free reaction this results from its ability to direct RNA polymerase II to the start site in cooperation with TFIIB (41). YY1 can activate transcription in some promoters where it does not bind at the start site (2,8,18,29,33–38); 124 (8,18); 25 (2); 129 (34); 110 (29); 136 (35). Our results indicate that activation versus repression does not depend on the specific YY1 sequence, since the wide range of binding sites tested in our experiments all elicited repression in our promoter constructs. Conceivably, the availability of factors that bind YY1 in different cell types influences its activity; a possibility supported by our confirmation of the observation that a GAL4–YY1 fusion protein will repress transcription in HeLa cells and activate the same construct in PYS-2 cells (35). Alternatively, additional factors

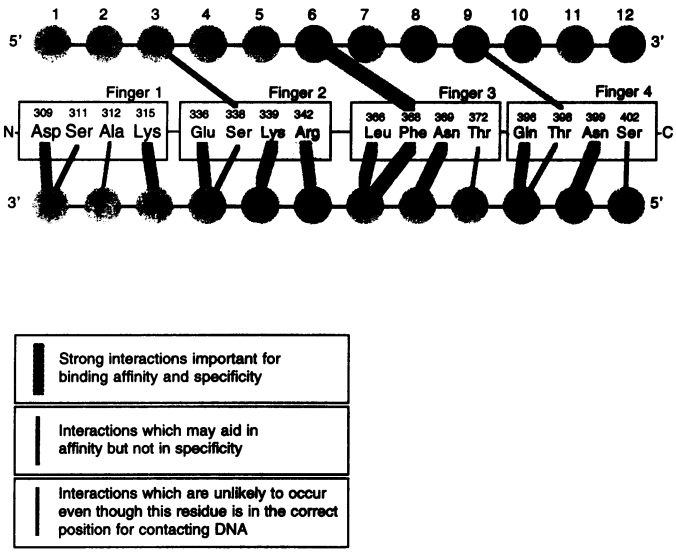

Figure 5. Predicted contacts between amino acid residues in the zinc fingers of YY1 with a consensus binding site. Bases comprising DNA strands are shown at the top and bottom of the diagram. The four zinc fingers are represented by the boxes and the numbers above each amino acid represent their positions in the human YY1 protein. Strong interactions, predicted to be important for specificity, are shown by bars with hatched lines, probable weaker interactions are depicted by thinner black bars and potential interactions unlikely to be important are designated by the grey bars.

interacting with the promoter at which YY1 is bound cause YY1 to activate rather than repress transcription.

Model for the YY1–DNA interaction

Three of the four zinc finger motifs in YY1 are members of the GLI family of zinc fingers and three-dimensional structures of protein–DNA co-crystals for several GLI family members have been solved. Analysis of these structures has suggested rules which predict the interactions that occur between amino acids and nucleotides when a zinc finger protein binds at its DNA recognition site (56–58). Suzuki *et al.* (58) have used these rules to generate interaction models for a variety of zinc finger domains, including the interaction of YY1 fingers 2–4 with a 9 bp recognition site. Our analysis of YY1 binding sites has revealed a conserved 9 bp binding motif (Tables 1 and 2). However, if it is assumed that all four zinc fingers of YY1 interact with DNA and each recognizes a 3 bp subsite (47,48), one would expect YY1 to interact with a 12 bp sequence. A tentative explanation for this apparent paradox and for the variability of YY1 binding sites presents itself upon consideration of the likely contacts in the major grove of the DNA helix with residues in the YY1 zinc finger motifs. These interactions are diagrammed in Figure 5. We will refer to bases comprising the bottom strand of DNA in Figure 5 in the following discussion, since it is the target of most YY1 contacts.

As discussed above, the only invariant sequence within the YY1 binding site is the central 5′-ATG-3′ (bases 8, 7 and 6 of the binding site, Fig. 5). This invariant element is near the middle of the YY1 binding interactions identified by DNase I footprinting and methylation interference (1,2,4,6,10,11,13,17,35,40,49). Since Arg has a strong preference for binding to G, particularly when the Arg is interacting at position 6[c] in the helix domain of the zinc finger (57,59), our model is anchored on an interaction between YY1 Arg342 (position 6[c] of zinc finger 2) and the G6 of the binding motif, an interaction originally proposed for this residue of finger 2 by Suzuki *et al.* (58). Given this anchor, a series of additional interactions can be predicted. Asn and Gln residues can bind to A (60,61) and show a strong preference for this interaction when Asn binds at position 3[b] of the helix (57,62). This recognition preference would be consistent with interactions of Asn399, Gln396 and Asn369 of YY1 at DNA positions A8, A10 and A11 respectively. The preference for an A > C at base 4 of the binding site can be explained by hydrogen bond formation with Glu336 of YY1. Lys has been shown to interact with G in zinc fingers of GLI, the glucocorticoid receptor, λ repressor and the *E.coli* Met J repressor (47,48), suggesting that YY1 Lys315 and Lys339 could interact with G3 and G5 respectively. YY1 clearly prefers base 7 of its recognition site to be a T, which is consistent with an interaction between Leu366 and the methyl group of thymine. In addition, Phe368 may contribute to the specificity for T at this position; this amino acid could also interact with C6 of the top DNA strand, providing an explanation for the strong G6 preference on the bottom strand (57,58). There is no preferred base at position 9 of the binding site, because YY1 Thr372 is potentially able to form hydrogen bond contacts with any of the bases, but is probably too small to even reach the DNA when located in position 6[c] of the zinc finger (56–58). The same logic also holds for position 12 of the binding site and Ser402 of YY1. YY1 Ala312 would be able to interact with T at position 2 of the binding site, but it is also not likely to contact the DNA, due to the small size of its side chain (56–58), explaining the flexibility at position 11 of the binding site. Finally, the heterogeneity seen at position 1 can be explained by Asp309 being too small to interact with DNA when placed in the –1[a] position of the finger.

This model for a 12 bp YY1 DNA binding site fits the rules that have been proposed for the interaction of GLI-type zinc fingers with DNA, it is consistent with DNase I footprinting and methylation interference studies and it can explain the considerable flexibility of the YY1 recognition site sequence at positions 1, 2, 9 and 12.

REFERENCES

1 Shi,Y., Seto,E., Chang,L.S. and Shenk,T. (1991) *Cell*, **67**, 377–388.
2 Hariharan,N., Kelley,D.E. and Perry,R.P. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9799–9803.
3 Flanagan,J.R., Becker,K.G., Ennist,D.L., Gleason,S.L., Driggers,P.H., Levi,B.-Z., Appella,E. and Ozato,K. (1992) *Mol. Cell. Biol.*, **12**, 38–44.
4 Park,K. and Atchison,M.L. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9804–9808.
5 Pisaneschi,G., Ceccotti,S., Falchetti,M.L., Fiumicino,S., Carnevali,F. and Beccari,E. (1994) *Biochem. Biophys. Res. Commun.*, **205**, 1236–1242.
6 Gualberto,A., LePage,D., Pons,G., Mader,S.L., Park,K., Atchison,M.L. and Walsh,K. (1992) *Mol. Cell. Biol.*, **12**, 4209–4214.
7 Lee,T. C., Shi,Y. and Schwartz,R.J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9814–9818.
8 Lee,T.C., Zhang,Y. and Schwartz,R.J. (1994) *Oncogene*, **9**, 1047–1052.
9 MacLellan,W.R., Lee,T.C., Schwartz,R.J. and Schneider,M.D. (1994) *J. Biol. Chem.*, **269**, 16754–16760.
10 Natesan,S. and Gilman,M.Z. (1993) *Genes Dev.*, **7**, 2497–2509.
11 Peters,B., Merezhinskaya,N., Diffley,J.F. and Noguchi,C.T. (1993) *J. Biol. Chem.*, **268**, 3430–3437.
12 Raich,N., Clegg,C.H., Grofti,J., Romeo,P.H. and Stamatoyannopoulos,G. (1995) *EMBO J.*, **14**, 801–809.

13 Yost,S.E., Shewchuk,B. and Hardison,R. (1993) *Mol. Cell. Biol.*, **13,** 5439–5449.
14 Ye,J., Ghosh,P., Cippitelli,M., Subleski,J., Hardy,K.J., Ortaldo,J.R. and Young,H.A. (1994) *J. Biol. Chem.*, **269,** 25728–25734.
15 Ye,J., Young,H.A., Ortaldo,J.R. and Ghosh,P. (1994) *Nucleic Acids Res.*, **22,** 5672–8.
16 Vincent,C.K., Gualberto,A., Patel,C.V. and Walsh,K. (1993) *Mol. Cell. Biol.*, **13,** 1264–1272.
17 Iannello,R.C., Kola,I. and Dahl,H.M. (1993) *J. Biol. Chem.*, **268,** 22581–22590.
18 Lee,Y.M. and Lee,S.C. (1994) *DNA Cell Biol.*, **13,** 1029–1036.
19 Lu,S.Y., Rodriguez,M. and Liao,W.S. (1994) *Mol. Cell. Biol.*, **14,** 6253–6263.
20 Gaston,K. and Fried,M. (1994) *FEBS Lett.*, **347,** 289–294.
21 Meier,V.S. and Groner,B. (1994) *Mol. Cell. Biol.*, **14,** 128–137.
22 Raught,B., Khursheed,B., Kazansky,A. and Rosen,J. (1994) *Mol. Cell. Biol.*, **14,** 1752–1763.
23 Bauknecht,T., Angel,P., Royer,H.D. and Zur,H.H. (1992) *EMBO J.*, **11,** 4607–4617.
24 May,M., Dong,X.P., Beyer-Finkler,E., Stubenrauch,F., Fuchs,P.G. and Pfister,H. (1994) *EMBO J.*, **13,** 1460–1466.
25 Pfister,H., May,M., Dong,X.P., Stubenrauch,F. and Fuchs,P.G. (1994) *J. Cell. Biochem.*, **S18C**, 226.
26 Montalvo,E.A., Shi,Y., Shenk,T.E. and Levine,A. (1991) *J. Virol.*, **65,** 3647–3655.
27 Liu,R., Baillie,J., Sissons,J.G. and Sinclair,J.H. (1994) *Nucleic Acids Res.*, **22,** 2453–2459.
28 Margolis,D.M., Somasundaran,M. and Green,M.R. (1994) *J. Virol.*, **68,** 905–910.
29 Momoeda,M., Kawase,M., Jane,S.M., Miyamura,K., Young,N.S. and Kajigaya,S. (1994) *J. Virol.*, **68,** 7159–7168.
30 Zock,C., Iselt,A. and Doerfler,W. (1993) *J. Virol.*, **67,** 682–693.
31 Lewis,B.A., Tullis,G., Seto,E., Horikoshi,N., Weinmann,R. and Shenk,T. (1995) *J. Virol.*, **69,** 1628–1636.
32 Yoganathan,T. and Sells,B.H. (1992) *Biochem. Cell Biol.*, **70,** 787–791.
33 Riggs,K.J., Saleque,S., Wong,K.K., Merrell,K.T., Lee,J.S., Shi,Y. and Calame,K. (1993) *Mol. Cell. Biol.*, **13,** 7487–7495.
34 Azizkhan,J.C., Jensen,D.E., Pierce,A.J. and Wade,M. (1993) *Crit. Rev. Eukaryote Gene Expression*, **3,** 229–254.
35 Satyamoorthy,K., Park,K., Atchison,M.L. and Howe,C.C. (1993) *Mol. Cell. Biol.*, **13,** 6621–6628.
36 Millette,R.L. and Mills,L.M. (1994) *J. Cell. Biochem.*,. **S18C**, 35.
37 Becker,K.G., Swergold,G.D., Ozato,K. and Thayer,R.E. (1993) *Hum. Mol. Genet.*, **2,** 1697–1702.
38 Chen,S., Mills,L., Perry,P., Riddle,S., Wobig,R., Lown,R. and Millette,R.L. (1992) *J. Virol.*, **66,** 4304–4314.
39 Seto,E., Shi,Y. and Shenk,T. (1991) *Nature*, **354,** 241–245.
40 Basu,A., Park,K., Atchison,M.L., Carter,R.S. and Avadhani,N.G. (1993) *J. Biol. Chem.*, **268,** 4188–4196.
41 Usheva,A. and Shenk,T. (1994) *Cell*, **76,** 1115–1121.
42 Shrivastava,A., Saleque,S., Kalpana,G.V., Artandi,S., Goff,S.P. and Calame,K. (1993) *Science*, **262,** 1889–1892.
43 Inouye,C.J. and Seto,E. (1994) *J. Biol. Chem.*, **269,** 6506–6510.
44 Lee,J.-S., Galvin,K.M., See,R.H., Eckner,R., Livingston,D., Moran,E. and Shi,Y. (1995) *Genes Dev.*, **9,** 1188–1198.
45 Lee,J.S., Galvin,K.M. and Shi,Y. (1993) *Proc. Natl. Acad. Sci. USA*, **90,** 6145–6149.
46 Seto,E., Lewis,B. and Shenk,T. (1993) *Nature*, **365,** 462–464.
47 Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61,** 1053–1095.
48 Pavletich,N.P. and Pabo,C.O. (1993) *Science*, **261,** 1701–1707.
49 Chung,S. and Perry,R.P. (1993) *Nucleic Acids Res.*, **21,** 3301–3308.
50 Bucher,P. (1993) 37 Ed, Postfach 10.2209, D-6900 Heidelberg, Germany.
51 Shrivastava,A. and Calame,K. (1994) *Nucleic Acids Res.*, **22,** 5151–5155.
52 Blackwell,T.K. and Weintraub,H. (1990) *Science*, **250,** 1104–1110.
53 Cantron,K.M., Iler,N. and Abate,C. (1993) *Mol. Cell. Biol.*, **13,** 2354–2365.
54 Pollock,R. and Treisman,R. (1990) *Nucleic Acids Res.*, **18,** 6197–6204.
55 Ekkar,S.C., Young,K.E., Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10,** 1179–1186.
56 Suzuki,M. (1994) *Structure*, **2**(4), 317–326.
57 Suzuki,M. and Yagi,N. (1994) *Proc. Natl. Acad. Sci. USA*, **91,** 12357–12361.
58 Suzuki,M., Gerstein,M. and Yagi,N. (1994) *Nucleic Acids Res.*, **22,** 3397–3405.
59 Berg,J.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89,** 11109–11110.
60 Gogos,J.A., Hsu,T., Bolton,J. and Kafatos,F. C. (1992) *Science*, **257,** 1951–1955.
61 Thukral,S.K., Morrison,J.L. and Young,E.T. (1992) *Mol. Cell. Biol.*, **12,** 2784–2792.
62 Fairall,L., Schwabe,J.W.R., Chapman,L., Finch,J. and Rhodes,D. (1993) *Nature*, **366,** 483–487.