*4700–4708* *Nucleic Acids Research, 1996, Vol. 24, No. 23* 

# Differential expression of the human ST5 gene in HeLa-fibroblast hybrid cell lines mediated by YY1: evidence that YY1 plays a part in tumor suppression

**Jack H. Lichy[1,2,*], Mourad Majidi[1], Jennifer Elbaum[1] and Mark M. Tsai[1]**

[1]Department of Cellular Pathology, Armed Forces Institute of Pathology, Washington, DC 20306-6000, USA and
[2]Laboratory of Tumor Virus Biology, National Cancer Institute, Bethesda, MD 20892, USA

Received August 19, 1996; Revised and Accepted October 9, 1996 DDBJ/EMBL/GenBank accession nos U15131, U15780, U15779, U15132

## ABSTRACT

**Through a mutational analysis of a differentially regulated enhancer, we present evidence that supports a role for the transcription factor YY1 in tumor suppression in HeLa/fibroblast somatic cell hybrids. The human ST5 gene was previously shown to be expressed as three RNA species, 4.6, 3.1 and 2.8 kb in length. Whereas the two larger species are expressed at similar levels in all cell lines examined, the 2.8 kb mRNA is expressed specifically in non-tumorigenic hybrids. In this study, the basis for the differential expression of this mRNA species was investigated. The message was shown to originate from a promoter located within an intron of the ST5 gene. An enhancer located ~1500 nt upstream of the start site was required for cell type specific expression. Mutational analysis of this enhancer revealed an AP1 site and five YY1 sites which were necessary for full enhancer activity. Levels of YY1 DNA binding activity were found to be as much as 6-fold higher in the non-tumorigenic cells relative to the tumorigenic cells, while AP1 activity was similar in both cell types. These results suggest that a signaling pathway targeting YY1 may play an important role in tumor suppression in HeLa–fibroblast hybrids.**

## INTRODUCTION

Somatic cell hybrids formed by fusion of HeLa cells to non-tumorigenic fibroblasts provide a model system for studying alterations in gene expression in carcinogenesis (1–3). The phenotypes of the hybrid cell lines derived from such fusions parallel that of cells at various stages of carcinogenic progression in the cervix, the anatomical site of origin of the HeLa cell line. In cervical carcinogenesis, expression of the transforming genes of a human papillomavirus (HPV) leads to abnormalities in cell growth control, but is not in itself sufficient to convert the benign keratinocyte into a carcinoma (4). Similarly, transfection of HPV DNA into primary cultures of human keratinocytes confers enhanced growth potential on the cells, but fails to produce a fully malignant phenotype (5). Such HPV 'transformed' keratinocytes grow with a flat morphology in culture and are non-tumorigenic upon injection into nude mice. In contrast, most cervical carcinoma cells prove to be tumorigenic in mice and grow with a characteristic transformed morphology in culture.

In the HeLa–fibroblast system, the hybrid cell lines initially demonstrate a non-tumorigenic phenotype. On prolonged passage these non-tumorigenic hybrids give rise to segregants which have regained the tumorigenic phenotype. In their morphology and lack of tumorigenicity, the non-tumorigenic hybrids resemble HPV transformed keratinocytes. In addition, these hybrids express the transforming genes of HPV18 at the same or slightly lower levels than the parental HeLa cell line (6). The tumorigenic segregants, in an analogous manner, show phenotypic similarities with cervical carcinoma cell lines. However, whereas cervical cancer cells might be expected to have undergone multiple genetic alterations relative to a benign keratinocyte, the tumorigenic segregant lines are more likely to differ from their non-tumorigenic parents predominantly in the expression of those genes required for the regulation of tumorigenicity (1,2,7).

The specific molecular events leading from the HPV transformed keratinocyte to carcinoma remain poorly understood. Some experimental evidence suggests that mutational activation of a *ras* oncogene can in some cases suffice to cause this phenotypic transition (8,9). Another series of studies, using the somatic cell hybrids as a model system, has led to the suggestion that the transformation results from alterations in a regulator of HPV transcription such that deregulated expression of the transforming proteins of HPV leads to the fully malignant phenotype (4,6,10).

In this study, we present evidence that at least some of the alterations in gene expression that distinguish the tumorigenic from the non-tumorigenic hybrids can be attributed to transcriptional regulation through YY1, a factor which has in fact been implicated as a regulator of HPV transcription. YY1 (also referred to as NF-E1, δ and UCRBP) is a bifunctional transcription factor which can act as either an activator or repressor of transcription depending on the specific context of its binding site. In addition to its effects on the long control regions (LCRs) of HPV16 and 18, YY1 has been implicated in the regulation of several cellular and viral regulatory regions [reviewed in (11)]. The YY1 protein contains distinct activator and repressor domains (12). In the regulatory region of HPV 18, YY1 functions either as an activator or repressor depending on the presence or absence of a 'switch' sequence (13). A role for YY1 in cervical carcinogenesis has been suggested by the observation of several cervical carcinomas in

*To whom correspondence should be addressed at: Department of Cellular Pathology, Armed Forces Institute of Pathology, 14 Street and Alaska Avenue, NW, Washington, DC 20306-6000, USA. Tel: +1 202 782 2562; Fax: +1 202 782 7623; Email:lichy@email.afip.osd.mil

which the HPV 16 promoter has escaped cellular repression via deletion of YY1 binding sites (14).

We previously reported the identification of a cellular gene which showed a striking pattern of differential regulation that correlated with tumorigenicity (15). To avoid ambiguity in nomenclature, this gene, which we previously named HTS1, will be designated by its locus name, ST5 (16,17). Non-tumorigenic hybrids express three ST5 RNA species, 4.6, 3.1 and 2.8 kb in length. The tumorigenic hybrids express similar or slightly reduced levels of the two larger species, but express very markedly reduced levels of the 2.8 kb transcript. This pattern of expression was of interest because the tumorigenic and non-tumorigenic hybrids are closely related genetically, and show extremely few differences in gene expression (3,18). The correlation between expression of the 2.8 kb transcript and the non-tumorigenic phenotype applied not only to the HeLa–fibroblast system but also to the non-tumorigenic HPV immortalized keratinocytes, which expressed the transcript, and to several cervical carcinoma cell lines, which showed markedly reduced levels of expression (15).

Regardless of whether the protein encoded by the 2.8 kb RNA plays a direct role in determining the non-tumorigenic phenotype, the pattern of expression of this transcript suggested that the mechanism responsible for its regulation might play an important role in the biology of tumor suppression in this system. In this report, we present the structure of the three full length ST5 mRNA species and characterize the mechanism of regulation of the 2.8 kb transcript. This RNA was found to be driven by a promoter located within an intron of the ST5 gene and to initiate near the 3′ splice site of this intron. Sequential deletions of this promoter demonstrated the presence of an enhancer ~1500 nt upstream of the initiation site. A mutational analysis of this enhancer revealed a cluster of five YY1 sites which play a critical role in mediating the specific expression of the 2.8 kb transcript in non-tumorigenic cell lines. Our results suggest that regulation of transcription by YY1 may be a mechanism of action of the tumor suppressor gene active in the HeLa–fibroblast hybrids.

## MATERIALS AND METHODS

### Extension of cDNA clones by 5′-RACE

Four successive applications of the 5′-RACE procedure (19) yielded 4277 nt of ST5 cDNA sequence (GenBank #U15131). This sequence was judged to be the full length or nearly full length 4.6 kb transcript by the following criteria: (i) 4277 is close to the length of 4.6 kb estimated from Northern blots; (ii) primer extension experiments with several oligonucleotides yielded similar banding patterns consistent with 5′ termini at or just beyond the first nucleotide in the sequence; and (iii) additional rounds of 5′-RACE failed to yield additional upstream sequence.

### Cell culture and transient expression assays

Cell lines were maintained in DMEM plus 10% fetal calf serum. Cells were plated out at $5 \times 10^5$ cells per 10 cm plate the day before transfection. Calcium phosphate precipitates (1.0 ml) were prepared containing 20 µg of the enhancer construct and 1 µg of pCMV-βgal to control for transfection efficiency. The transfected cells were subjected to a 1 min glycerol shock 4 h after adding the precipitate. Extracts were prepared 72 h post transfection by three freeze-thaw cycles in 150 µl 0.25 M Tris, pH 8.0. Acetylation assays were carried out in a final volume of 75 µl that included 20 µl lysate (50–70 µg protein), 0.1 µCi [$^{14}$C]chloramphenicol (Amersham), 0.25 M Tris, pH 8.0, and 1 mM acetyl-CoA. Except where indicated otherwise, reactions were incubated for 2 h at 37°C, extracted with ethyl acetate, and analyzed by thin layer chromatography in 96% chloroform/4% methanol. Reaction products were quantitated on a Molecular Dynamics PhosphorImager and percent acetylation was standardized to β-galactosidase activity in the same extract. For each set of CAT assays presented, the activity indicated for the wild type construct represents the actual %-acetylation observed. The tumorigenic and non-tumorigenic hybrids used in this study demonstrated very similar transfection efficiencies, as assessed by quantitation of β-galactosidase activity in parallel transfections of the two cell lines.

### Library screening

The partial genomic clone of the ST5 gene was isolated from a Charon 4A human genomic library (21) purchased from ATCC (#37333).

### Primer extension

$^{32}$P-end labeled oligonucleotide primer (1 ng) was mixed with 5 µg polyadenylated RNA in 18 µl reactions containing 0.5 mM dNTPs and RT buffer (GIBCO/BRL). After incubation at 95°C for 3 min and 50°C for 60 min, 200 U of SuperScript reverse transcriptase (GIBCO/BRL) was added and the incubation was continued for an additional 60 min at 50°C. After phenol extraction and ethanol precipitation, the reaction products were separated on 6% acrylamide/7 M urea sequencing gels and detected by exposure to a PhosphorImager screen (Molecular Dynamics).

### Nuclease protection

The probe used for S1 protection assays was an oligonucleotide 60 bp in length complementary to nt 1579–1638 of the cDNA sequence. Hybridizations were performed at 50°C for 2 h in 20 µl reactions containing 1 ng of the $^{32}$P-end labeled oligonucleotide, 20 µg cytoplasmic RNA from the indicated cell line, 1 M NaCl, 1 mM EDTA and 133 mM HEPES, pH 7.2. S1 digestion was carried out at 13°C for 2 h by bringing the reaction volume to 150 µl with 0.25 M NaCl; 30 mM sodium acetate, pH 4.5; 1 mM $ZnSO_4$ and adding 400 U S1 nuclease (Boehringer-Mannheim). Reactions were stopped by the addition of EDTA to 10 mM followed by phenol/chloroform extraction and ethanol precipitation. Products were analyzed by electrophoresis on an 8% sequencing gel. Bands were visualized with a Molecular Dynamics PhosphorImager.

### Mutant construction

Unidirectional Exonuclease III deletions were constructed by using the Exo/Mung kit (Stratagene) according to the manufacturer's instructions. The panel of clustered point mutations in the 85 nt enhancer fragment was constructed by reconstruction of this region with three overlapping pairs of oligonucleotides. The structure of each mutant was confirmed by DNA sequencing with the Sequenase kit (Amersham).

### Electrophoretic mobility shift (EMSA) assays

Nuclear extracts used for electrophoretic mobility shift assays (EMSA) were prepared by the method of Dignam *et al.* (22).

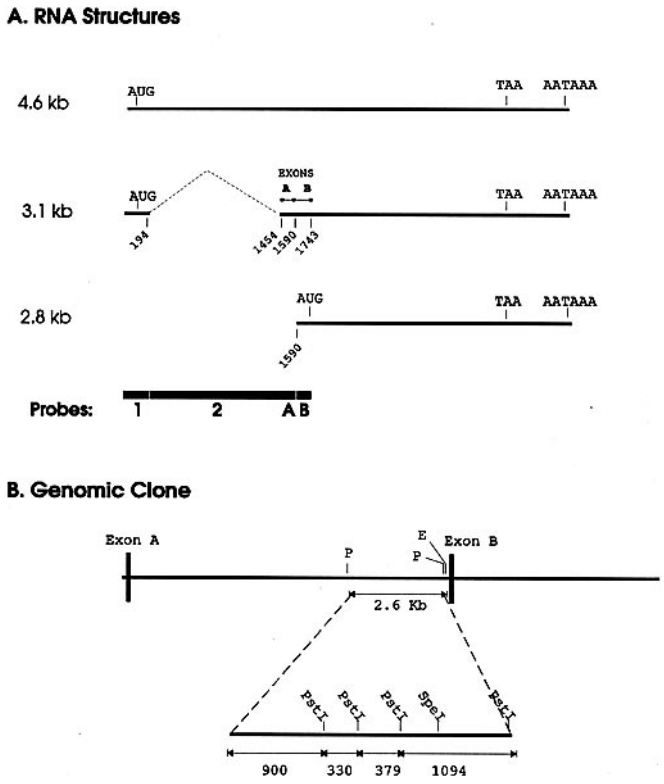


**Figure 1.** Structure of ST5 RNA species and a partial genomic clone. (**A**) Structure of the three ST5 RNA species. The designations '4.6,' '3.1' and '2.8' kb refer to the apparent length of these transcripts on Northern blots. (**B**) Structure of the genomic insert from a bacteriophage λ clone. The location of the 2.6 kb *Pvu*II fragment within the 9 kb intron separating exons A and B is shown. P: *Pvu*II; E: *Eco*RI.

Reaction mixtures (15 µl) contained 1 ng of $^{32}$P-end labeled probe, 1.5 µg poly d(I–C) and 4 µg of protein in 10 mM Tris, pH 7.5, 4% glycerol, 1 mM EDTA and 150 mM NaCl (final concentration). Reactions were incubated at 4°C for 15 min, loaded onto a pre-run 4% polyacrylamide gel in 0.25× TBE, and run at 250 V and 4°C until bromophenol blue run in a separate lane was 1–2 cm from the bottom of the gel. The competitors were all double stranded oligonucleotides, and were added to the reactions where indicated prior to the addition of nuclear extract.

### GenBank accession numbers

The sequences of the 4.6, 3.1 and 2.8 kb ST5 cDNAs have been assigned GenBank accession numbers U15131, U15780 and U15779. 2.3 kb of sequence from the promoter containing intron extending upstream from the splice junction and containing the enhancer sequence has been submitted to GenBank with the accession number U15132.

## RESULTS

### Structure of the ST5 RNA species

The sequence of the full length ST5 cDNA, corresponding to the 4.6 kb transcript detected on Northern blots, was obtained as described in Materials and Methods. A series of RT–PCR and Northern blot experiments were then performed to determine the

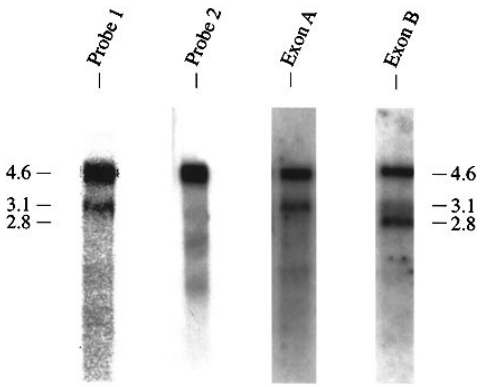


**Figure 2.** Northern blot analysis of RNA from the non-tumorigenic hybrid ESH 5E with ST5 cDNA probes. Polyadenylated RNA (10 µg) was separated in four lanes of a 1.2% agarose/6% formaldehyde gel and transferred to a nylon membrane. Strips from this gel were hybridized to probes consisting of segments of the ST5 cDNA sequence, shown schematically in Figure 1. The segments consisted of the following bases in the ST5 cDNA sequence (numbered as in GenBank #U15131): Probe 1: 1–194; Probe 2: 195–1454; A: 1476–1590; B: 1592–1710. Blots were washed and exposed for 1–7 days to either X-ray film or a PhosphorImager screen.

relationship of the smaller two RNA species to the 4.6 kb message. The results support the structures shown in Figure 1A. To identify alternatively spliced species, RT–PCR reactions were carried out with an upstream primer containing sequence near the 5′ end of the message and a downstream primer from one of several regions within the cDNA. With a downstream primer taken from within the 3′-terminal 2823 nt, these PCR reactions produced two products, the larger one of which had the size and sequence expected from the full length cDNA sequence. Sequencing of the smaller product revealed a deletion of nt 195–1454, suggesting the existence of an alternatively spliced product lacking this region. Consistent with this interpretation, a probe containing only this region recognized exclusively the 4.6 kb ST5 message on a Northern blot (Probe 2 in Fig. 2). A probe upstream of this region (Probe 1) recognized both the 4.6 and 3.1 kb RNA species but not the 2.8. Several probes tested from the downstream 2687 nt recognized all three species. These results indicated that the 4.6 and 3.1 kb ST5 transcripts represent alternatively spliced mRNAs with common 5′ and 3′ ends.

### Transcription of the 2.8 kb ST5 RNA initiates at an internal promoter

The absence of the Probe 1 sequence in the 2.8 kb transcript, together with the lack of evidence for an alternative 5′ exon after sequencing multiple RACE products, suggested that this transcript might be driven by a promoter located internally in the gene. To investigate this possibility, a probe consisting of nt 1593–1958 [the most upstream portion of the previously reported sequence (15)], which recognized the 2.8 kb species on a Northern blot, was used to isolate a partial genomic clone of the ST5 gene. Two positive phage λ clones were isolated, one of which contained 9.5 kb of genomic sequence upstream of the *Bgl*II site at nt 1593. By sequencing, restriction mapping and hybridization to cDNA probes, this clone was found to contain two exons of the ST5 gene, designated exon A and exon B (Fig. 1B). On Northern blots (Fig. 2), a probe consisting of the exon B sequence recognized all

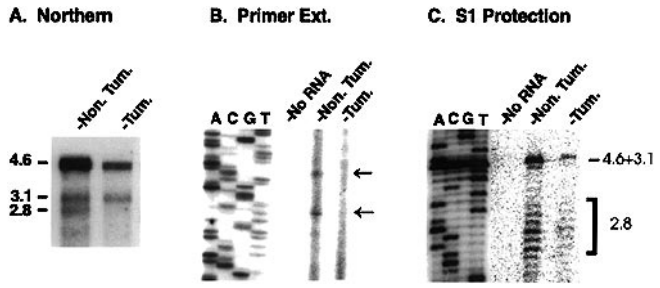


**Figure 3.** Primer extension and nuclease protection analysis of ST5 RNA. (**A**) Northern blot analysis of RNA from the non-tumorigenic HeLa–fibroblast hybrid ESH 5E and the tumorigenic segregant ESH 5L, illustrating the relative levels of expression of the three ST5 transcripts. (**B**) Primer extension assay. Primer extension was performed as described in Materials and Methods with an oligonucleotide containing the ST5 sequence complementary to nt 1751 to 1722 in Figure 1. (**C**) S1 Protection. RNA from the indicated cell lines was hybridized to an antisense 60 base oligonucleotide probe complementary to the cDNA sequence and spanning the splice junction. Hybridization and S1 digestion were carried out as described in Materials and Methods. In B and C, the sequencing ladders shown were run on the same gel as the primer extension or S1 products, and were aligned as shown in the figure. To generate the figures, the reaction products and markers were imaged separately and the grey scale was adjusted for optimal visualization of bands on each image. The two images were then realigned to create the images shown.

three ST5 RNAs. In contrast, the 136 nt exon A sequence detected the 4.6 and 3.1 kb species but did not recognize the 2.8 kb RNA. These results suggested that the 5′ end of the 2.8 kb message was within or immediately upstream of exon B.

The localization of the 5′ end was defined more precisely with the primer extension and nuclease protection experiments presented in Figure 3B and C. The relative levels of the three mRNA species in tumorigenic and non-tumorigenic hybrids are illustrated in the Northern blot shown in Figure 3A. For primer extension (Fig. 3B), a 30 nt antisense oligonucleotide primer yielded major primer extension products mapping putative start sites to positions –2 and +15 relative to the 3′ splice junction. Neither band was detected in the absence of RNA (lane 1), or with RNA from the tumorigenic hybrid (lane 3). These results parallel the differential expression of the 2.8 kb RNA in the tumorigenic and non-tumorigenic hybrids (Fig. 3A), and support the conclusion that the bands detected correspond to the 5′ termini of this RNA species. In the nuclease protection assay (Fig. 3C), an antisense oligonucleotide probe complementary to nt 1579–1638 of the cDNA, and spanning the splice junction at nt 1590–1591, was incubated either with no RNA or with RNA from the tumorigenic or non-tumorigenic hybrids followed by digestion with nuclease S1. A band corresponding to protection of the full length oligonucleotide probe, resulting from the 4.6 and 3.1 kb transcripts, was detected with somewhat greater intensity with RNA from the non-tumorigenic hybrid, consistent with the expression pattern of these transcripts (Fig. 3A). In contrast, a cluster of digestion products centered 2–3 nt upstream of the splice site showed marked specificity for the non-tumorigenic cell line, as expected for the 2.8 kb transcript. The primer extension and S1 protection data argue for a major initiation site of the 2.8 kb RNA 2–3 nt upstream of the splice junction.

## The intron upstream of exon B contains a promoter specific for non-tumorigenic hybrids

The localization of the initiation site for the 2.8 kb mRNA suggested that the promoter driving transcription of this message might lie within the intron between exons A and B. To test this possibility, a 2.6 kb *Pvu*II fragment containing the downstream portion of this intron (Fig. 1) was cloned into pSV0CAT. The 2.6 kb *Pvu*II fragment drove expression of the CAT gene in transient expression assays in an orientation dependent manner (Fig. 4). Furthermore, this activity demonstrated a marked specificity for the non-tumorigenic relative to the tumorigenic HeLa–fibroblast hybrid. Control expression plasmids, including pSV2CAT, an HPV18 LCR-CAT construct and pCMV-βgal, the plasmid used for standardization of the CAT data, resulted in similar levels of CAT activity in the two cell lines. These results, together with the primer extension, nuclease protection, and Northern blot data, strongly support the conclusion that the promoter for the 2.8 kb transcript lies within the intron between exons A and B. Furthermore, since cell type specific regulation can be mediated by the untranscribed upstream sequence, the results suggest that regulation of this transcript occurs at the level of transcription.

## Structure of the internal promoter

A series of deletions were introduced into the 2.6 kb *Pvu*II promoter fragment to identify elements required for activity. The sequences of the regions found to be contributory to promoter activity are given in Figure 5. Deletions were generated upstream and downstream of a unique *Spe*I site located 836 nt upstream of the splice junction by exonuclease III digestion. Deletions upstream of the *Spe*I site were either without effect or resulted in a modest stimulation of activity (with no loss of specificity) until nt –1550 relative to the splice junction (Fig. 4). Deletions extending upstream of this point yielded markedly reduced activity. These deletion plasmids thus suggested that a region located upstream of nt –1550 is required for maximal expression in the non-tumorigenic hybrid cell line.

Deletions downstream of this site to a point just upstream of a G-rich region (italics in Fig. 5B) had no effect on either basal promoter activity or specificity for non-tumorigenic cells. Extension of the deletion through the G-rich region resulted in complete loss of promoter activity. An additional construct, labeled Ext 1(1598), extended the promoter fragment past the downstream *Pvu*II site to the splice junction. This plasmid demonstrated a marked enhancement of promoter activity. These results indicate that elements necessary for optimal promoter activity include the G-rich region, sequence between the downstream *Pvu*II site and the splice junction, and an element upstream of nt –1550.

## Specificity maps to an upstream enhancer region

To determine whether the property of specific expression in the non-tumorigenic cells mapped to any of the regions critical for activity, portions of the promoter were cloned in front of the enhancerless SV40 promoter in the plasmid pA10CAT and transiently transfected into tumorigenic and non-tumorigenic HeLa–fibroblast hybrids (data not shown). The promoter proximal region containing the G-rich region had no detectable effect on the basal activity of the vector in either cell line. A *Pst*I fragment containing nt –1497 to –1826 resulted in a 20–30-fold stimulation of CAT activity which was specific for the non-tumorigenic cell line and independent of the orientation of the fragment. Other

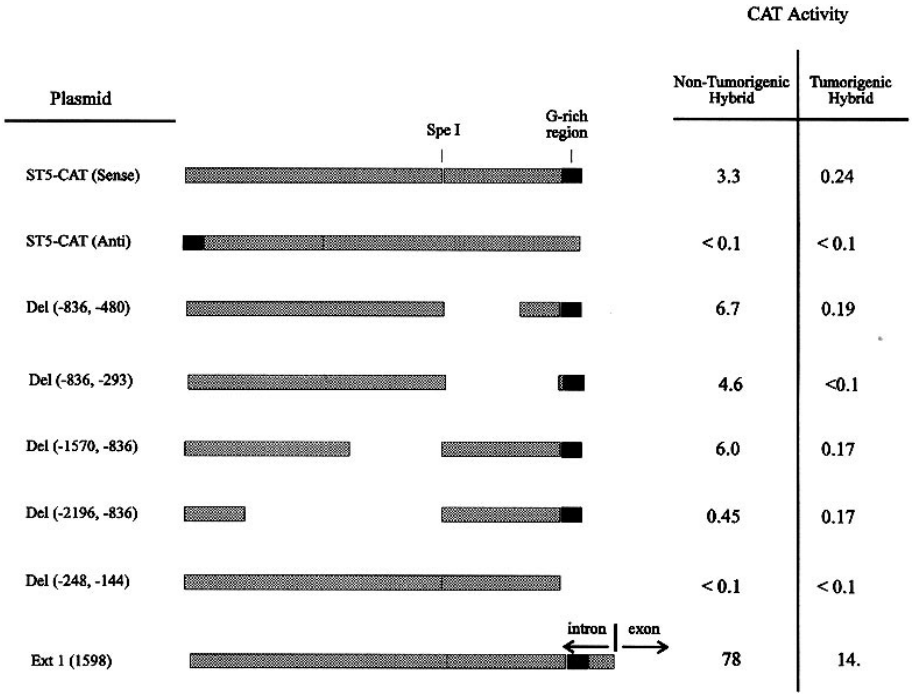


**Figure 4.** Transient expression analysis of the 2.6 kb promoter fragment and derivatives. The 2.6 kb *Pvu*II fragment was cloned into pSV0CAT in both orientations. Derivatives were prepared and transfected as described in Materials and Methods. Each plasmid was transfected at least three times in at least two independent experiments. CAT assays were incubated for 4 h. The data shown are the results of a typical experiment. For the deletion plasmids, the endpoints of the deletions are given in parentheses relative to the splice junction.

**A.**

```
1826 CTGCAGGGCATCTGTTGGGGAAGGATCAAGAAGTAGAGAATCCATCCCCAACACCTCCCT
1766 AAGTGCCCCACAGTGGGGTAGGCTCCAGAAAAGATGCAGGGGGTGCCGGGTTTGCTATTT
1706 GGCTTCTTGCGACAGGGCTGCTGGGCCACGTCCGCTTAGTCAGGGCTCTGGAATGGAGAC
1646 AAGTCATTTTAATAGCATGACTGTGTTACAGGAAAAGGTCTGTTCCATTGCTGGGGACTC
1586 TGCTTGGAGTTGGGAATCACTTAATCCAGTCCCAAATCCCAGTGACCCATCTTTGACTCA
1526 TAAGTTTACAGAGATTTATATTGCTCCCACTGCAG
```

**B.**

```
288  GTGCAATCCCTGAATTTATAAAGCAGAGGCCAGGCTGAGCTGGGCTGGGGGTGGGGGGTT
228  GGGGGGTGGGGGGGGGTGGGGGGGTGGGCTCTGCTTCTGGACTTGGAGCCAGGATCTGTT
168  TGTAGCCTCCCCACTGCAGACGGGCAGCTGCATGACTCATGAGCTGAGGGGCCAGGCAGA
108  AGCCATTCTCTTTGGTTCACAGGAATTCTGTTGTTTATTTGTGTGTGTGTGAGTTCGTGT
48   GTTTCCTTCTTCCCTTCTCCCCCTTCTCTAACTCTTCCTCCTCTGCAG GAGATCTGCCA
                                          ←intron exon→
```

**Figure 5.** Sequence of critical regions of the intronic promoter. Nucleotides are numbered from the junction between the intron and exon B, counting the first residue upstream of the junction as '–1.' (**A**) Sequence of the 330 nt. *Pst*I fragment containing enhancer activity. The 85 nt segment studied with clustered point mutations is indicated in bold italic type. (**B**) Sequence of proximal promoter and splice junction. The G-rich region is indicated in italics. The *Pvu*II site used as the downstream terminus of the 2.6 kb promoter fragment is in bold type.

fragments derived from the promoter sequence had no effect on promoter activity. Specific expression of the internal promoter therefore maps to an enhancer element located between positions –1497 and –1826 relative to the splice junction (Fig. 5A).

To localize specific elements required for activity, exonuclease III deletions were made from either end of this 330 nt fragment and tested in the transient expression assay. Deletions from –1826 to –1681 had no effect on enhancer activity. The results of additional deletions with schematics showing the AP1 and YY1 sites subsequently demonstrated to be important (as shown below) are given in Figure 6. Deletions downstream from nt –1681 resulted in an ~20-fold reduction in activity with deletions as small as 17 nt, which eliminated the AP1 site. Deletions from the opposite end of the fragment resulted in a more gradual loss of activity. Fifty percent of the activity remained with a deletion extending to a point just downstream of the YY1 cluster. Further deletion of the region containing four YY1 binding sites led to an additional 10-fold loss of activity, and deletion of all five YY1 sites reduced the activity 2–3-fold further.

A series of 20 clustered point mutations scanning the region from –1597 to –1681 was constructed to define more precisely the location of elements required for enhancer activity. The results of transient expression assays with these mutants are presented in Figure 7. These results are consistent with a requirement for multiple sites for full activity. Mutants 1 and 2 had similar activity to the wild type enhancer fragment. Mutants 3, 4 and 5, which involved parts of the AP1 consensus, demonstrated a very marked reduction in activity. Mutant 6, although not affecting a known factor binding site, also gave consistently low activity. Mutant 7, which disrupted the first YY1 site, had ~60–80% less activity than wild type. Following in the same pattern, mutations in the next three YY1 sites (mutants 10;13 and 14; 17 and 18) showed similarly low levels of activity while mutations not affecting a YY1 site had activities equal to or greater than wild type. Mutations in the fifth YY1 site (mutants 19 and 20) reproducibly resulted in a 1.5–2-fold increase in activity. Since YY1 can function in either a positive or negative manner, this result is also consistent with a role for YY1, although involvement of a distinct negative regulatory factor at this site cannot be excluded.

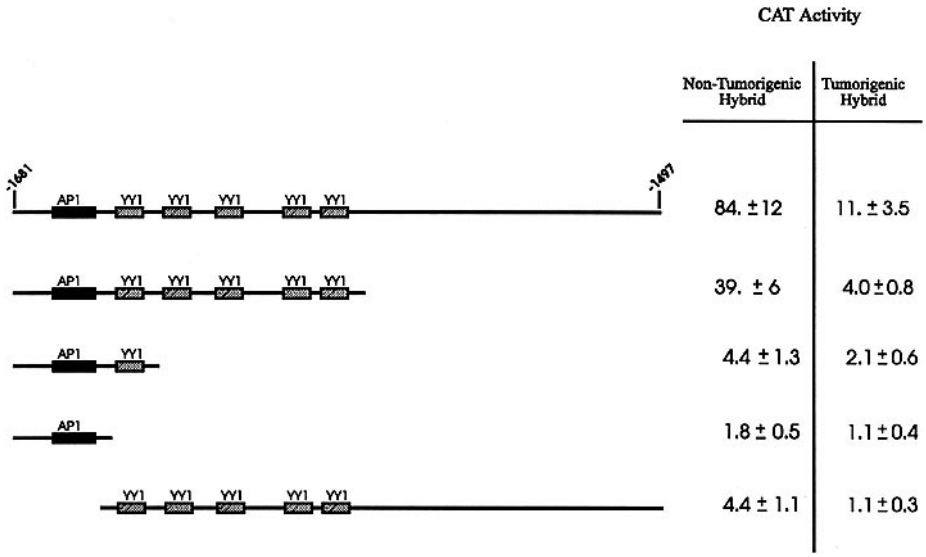


**Figure 6.** Exonuclease III deletions of 330 nt enhancer fragment. The 330 nt *Pst*I fragment (Fig. 4A) was cloned into the enhancerless reporter plasmid pA10CAT. Unidirectional Exonuclease III deletions were generated from each end of the insert by using the Exo/Mung kit (Stratagene). Deletion constructs were assayed for enhancer activity by transfection into the non-tumorigenic hybrid ESH 5E and the tumorigenic segregant ESH 5L. The diagram indicates the location of sequences subsequently identified as AP1 and YY1 sites. CAT activities shown represent the mean ± S.D. of three separate determinations.

## Identification of factors involved in specific expression

Several probes derived from the critical regions of the enhancer were employed in mobility shift assays with nuclear extracts from the tumorigenic and non-tumorigenic hybrids. A probe containing the AP1 consensus sequence demonstrated a prominent shifted band, present with approximately equal intensity with extracts from either cell line (Fig. 8A). This band was efficiently competed by an excess of the probe or an AP1 consensus competitor, but was not competed appreciably by the probe sequence with a mutation in the AP1 consensus.

A second probe, spanning the regions contained in mutants 6, 7 and 8, yielded a characteristic banding pattern with four prominent shifted bands (Fig. 8B). Complexes C1, C3 and C4 were generally present at similar levels with extracts of both cell types. Complex C2, however, was markedly more intense with extracts from the non-tumorigenic cell line. We therefore hypothesized that the factor(s) present in this complex might play an important role in the differential expression of the ST5 2.8 kb transcript.

The nucleotide sequences required for the formation of complex C2 were determined by gel shift competition assays (Fig. 9). A series of clustered point mutations scanning the probe sequence showed that only mutations in the sequence AATGGA (competitors 5 and 6 in Fig. 9) failed to compete for complex C2. Competitor 7 also appeared to compete with reduced efficiency. Single base mutations of the nucleotides flanking this region further demonstrated the importance of the first 'A' in the sequence ATGGA (competitor 9). Point mutants having alterations in the sequence downstream of this site (competitors 10, 11 and 12) competed as well as the control for complex C2, but with reduced efficiency for complex C3. All competitors tested produced approximately equal reductions in the intensity of complex C1 and generally had little effect on C4, indicating that these represent non-specific DNA-protein complexes. Complex C3 showed sequence specificity similar but not identical to C2.

To assess the relative roles of the factors represented by complexes C2 and C3, the mutations present in competitors 9 and 10 were introduced into the CAT reporter vector as mutants 21 and 22, respectively, and tested in the transient expression assay (Fig. 7). The activity of mutant 21, which could form C3 but not C2, was substantially reduced. In contrast, the activity of mutant 22, which could form C2 but not C3, was similar to that of the wild type enhancer. The results obtained with mutants 21 and 22, together with the analysis of the corresponding mutants in the mobility shift assays, implicate the factor involved in complex C2 as a positive regulator of transcription, and argue against a role for the C3 protein(s) as either a positive or negative regulator of transcription.

The sequence requirements for the formation of complex C2 were further characterized by mutating each nucleotide in the binding site ATGGAG inferred from the experiment in Figure 9 to every other nucleotide. Each mutant was tested as a competitor in the mobility shift assay (Fig. 10). This experiment revealed a binding consensus sequence of A-T-G-G/a-C/a/t/g, where a capital letter indicates the preferred nucleotide.

The derived consensus was similar to that reported for YY1 (11,23). The supershift assay shown in Figure 11 was performed to confirm the presence of YY1 in complex C2. This experiment shows that incubation with a YY1 antibody resulted in the loss of complex C2 in conjunction with the appearance of a supershifted complex. Neither band C1 nor C3 were affected by the YY1 antibody, nor were any of these complexes affected by antibodies to c-jun, c-fos, or the ets family of transcription factors, included as controls.

Each of the additional YY1 sites noted in Figure 7 were confirmed as such by competition experiments using oligonucleotides containing each of these sites as competitors for complex C2 (data not shown). Of these sites, the sequence AGGTC at nt –1611 is unusual in that it varies from the consensus sequence at what otherwise seems to be a critical nucleotide. Nevertheless, this

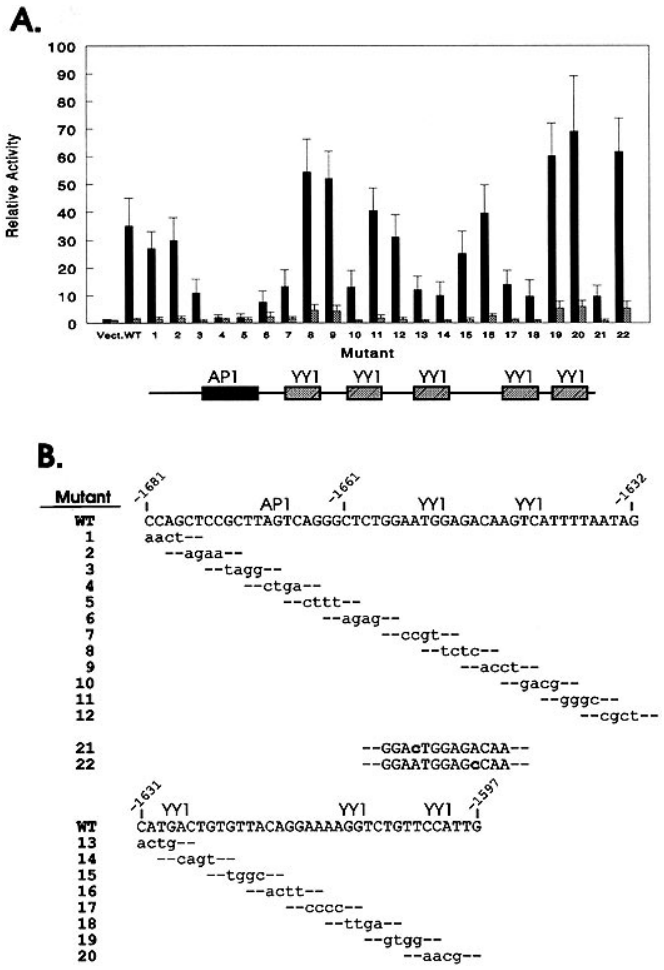


**Figure 7.** Mutagenic analysis of 85 nt enhancer segment. Clustered point mutations were introduced throughout the 85 nt region that conferred cell type specific enhancer activity. Each mutant was assayed for activity in the non-tumorigenic hybrid ESH 5E and the tumorigenic segregant ESH 5L. (**A**) Enhancer activity of mutants. Values shown represent the mean ± S.D. of three separate determinations. The diagram beneath the graph shows the location of the AP1 and YY1 binding sites in alignment with the corresponding mutants involving these sites on the X-axis. Solid and shaded bars represent results obtained with the non-tumorigenic hybrid and the tumorigenic segregant, respectively. The results shown are a composite of several experiments normalized to the activity of the wild type enhancer fragment in each individual experiment. (**B**) Sequence of enhancer mutants. Mutants 1–20 are clustered point mutations scanning the 85 nt enhancer segment. Mutants 21 and 22 are point mutations generated to draw correlations between bands detected on EMSA assays and enhancer activity.

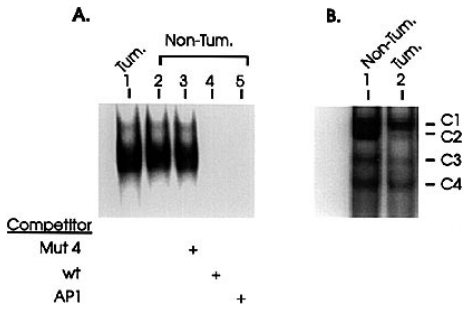


**Figure 8.** EMSA of critical regions in enhancer. EMSA assays were carried out under the conditions described in Materials and Methods. (**A**) Probe containing AP1 consensus. Nuclear extracts from either the non-tumorigenic ESH 5E (Non-tum.) or the tumorigenic segregant ESH 5L (Tum.) were incubated with 1 ng of an AP1 consensus probe (Pharmacia) end labeled with $^{32}$P. Equal amounts (4 μg) of protein were used to compare the two cell lines. Where indicated, 250 ng of competitor oligonucleotides were added to the reaction. The Mutant 4 competitor contained nt –1651 to –1681 with the AP1 site mutation contained in Mutant 4 (Fig. 7B). (**B**) The probe consisted of nt –1644 to –1667, encompassing the first YY1 consensus downstream of the AP1 site. The major mobility shifted bands are labeled C1–C4. In both (A) and (B) only the region of the autoradiogram containing the shifted bands is shown.

sequence competed as effectively for complex C2 as the other YY1 sites present in this region.

## DISCUSSION

We have studied the mechanism of activation of the ST5 2.8 kb transcript in non-tumorigenic cells and have obtained evidence suggesting a critical role for the transcription factor YY1. Although this factor has been implicated in the regulation of several genes, we believe that this is the first study to suggest a specific role for YY1 as a mediator of tumor suppression. The initial studies to characterize the 2.8 kb transcript involved defining the initiation site and obtaining a genomic clone of the promoter. Our results demonstrate that this transcript is driven by a promoter located within an intron of the ST5 gene, with transcriptional initiation occurring very close to the 3′ splice junction. The sequence upstream of the splice site contained a G-rich region, which was shown to be required for promoter activity. Although no perfect TATAA motif was present in the –30 region, the sequence TCTAA was present at that location, and potential TATAA motifs were noted further upstream. Three regions required for promoter activity were identified: the G-rich region, the sequence between this region and the start site, and an enhancer element located ~1500 nt upstream of the start site. Each functional region of the promoter was tested separately for the ability to confer cell type specific expression on a heterologous promoter. These experiments localized this activity to the upstream enhancer element.

A combination of deletion and clustered point mutations were employed to localize the specific sequences required for activity. Each region was then tested in gel mobility shift assays to assess differences in levels of DNA binding activities between tumorigenic and non-tumorigenic cells. These experiments led to two principal conclusions. First, the putative AP1 motif within the enhancer binds to similar factors as a consensus AP1 competitor, and similar levels of these factors are present in the nuclear extracts of tumorigenic and non-tumorigenic cells. Second, substantially (3–6-fold) higher levels of a DNA–protein complex containing the transcription factor YY1 are formed with nuclear extracts of the non-tumorigenic than the tumorigenic cell line. Given the initial observation that one of the motifs found to be necessary for enhancer activity bound YY1, it became apparent that four additional sites that led to altered enhancer activity when mutated also represented YY1 motifs. The ability of each of these sites to compete for YY1 binding was confirmed experimentally. Therefore, our results indicate the presence of one AP1 and five YY1 motifs important for enhancer activity. Four of the YY1 motifs have a positive effect on transcription, while the fifth functioned as a negative regulator in the context of our reporter constructs. Mutation of any one of the four positive regulatory motifs led to a 2–4-fold reduction in enhancer activity. Deletion of the entire

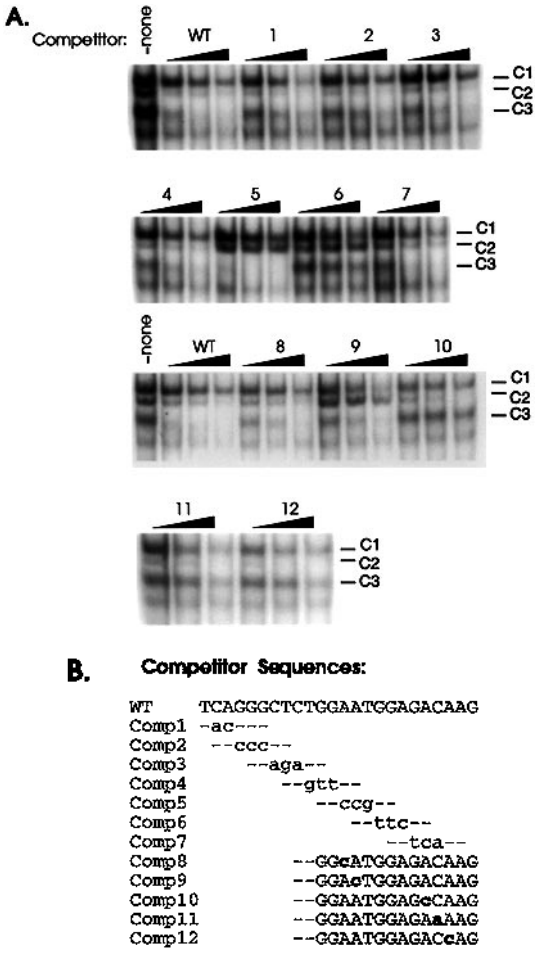


**Figure 9.** Localization of binding site for complex C2. EMSA assays were performed as in Figure 8B. The sequence of the wild type and mutant competitor oligonucleotides is given in (B). Each competitor was tested at 50, 100 and 250 ng per reaction, as indicated by the wedges above groups of three lanes.

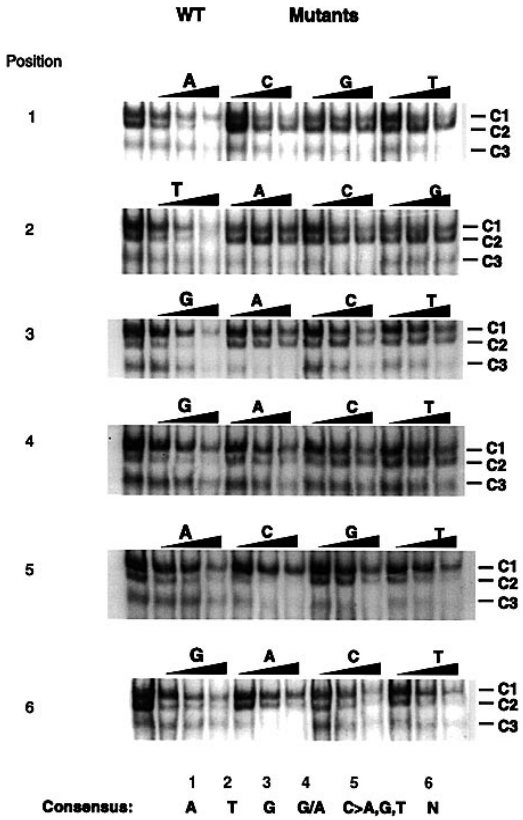


**Figure 10.** Sequence requirements for complex C2. EMSA assays were performed under the same conditions as in Figure 8B. A series of competitor oligonucleotides was tested, each of which contained a point mutation altering one of the bases in the deduced binding sequence for complex C2. In each grouping the first lane shows the banding pattern with no competitor added; the next three lanes show the effect of competition by the wild type oligonucleotide (identical to the probe).The remaining lanes show the effect of altering the nucleotide at the indicated position to each of the three other possible nucleotides.

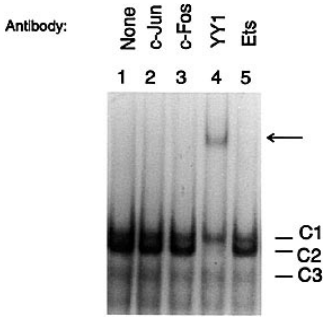


**Figure 11.** Detection of YY1 in complex C2 by supershift with a YY1 antibody. EMSA assays were performed under the same conditions as in Figure 8B except that 50 ng of the competitor 6 oligonucleotide pair was added in order to enhance complex C2 relative to the other bands. Nuclear extracts were preincubated with 1 µg of antibody to c-jun, c-fos, c-ets or YY1 (all purchased as 'Supershift' reagents from Santa Cruz Biotechnology) for 60 min at 0°C prior to being added to the EMSA reaction mixture.

sequence block containing the five YY1 motifs resulted in an ~20-fold loss of enhancer activity.

The YY1 binding site shows a high level of degeneracy (11,23). As a result, such sites are often difficult to recognize based on sequence alone. The identification of a DNA binding protein in complex C2 as YY1 required a detailed mutagenic analysis of the binding site. The binding requirements elucidated in the EMSA experiments closely match the consensus sequence 5′-SKCCATN-TT-3′ determined for YY1 (23). In this consensus, only the CAT core shows a high degree of conservation. Our analysis (Fig. 10) similarly shows a strong preference for this core sequence. Substitutions in the surrounding nucleotides had less pronounced effects on binding. In view of this consensus, it is interesting that one site shown to bind YY1 deviated from the consensus within the CAT core, but nevertheless competed for YY1 binding as well as a consensus competitor. This site was initially identified by the observation that Mutant 17 consistently gave reduced levels of CAT activity in transient expression assays (Fig. 7).

The finding that YY1 level, as assessed by DNA binding activity, varies substantially between non-tumorigenic and tumorigenic HeLa–fibroblast hybrids, is of interest in the context of studies of HPV gene regulation in these cell lines. Several recent studies point to an important role for YY1 in the regulation of HPV transcription. The long control region (LCR) of HPV16 contains

six YY1 and two AP1 sites involved in transcriptional regulation (14). At least some of the YY1 sites appear to function as negative regulatory elements which reduce expression of the transforming genes of the virus. Analysis of cervical carcinomas containing episomal HPV16 revealed two cases in which the viral genome had undergone deletions of several of these YY1 sites (14). These deletions resulted in increased transcriptional activity, suggesting that escape from YY1 inhibition might be a mechanism of carcinogenic progression in these tumors. The HeLa–fibroblast hybrids express the transforming genes of HPV18. It might therefore be expected that the elevated levels of YY1 detected in the non-tumorigenic hybrids would be reflected in the level of HPV18 RNA expression in these cell lines. In conflict with this expectation is the observation that the tumorigenic and non-tumorigenic hybrids express similar levels of viral RNA (6). Rosl (10) has reported evidence that the tumor suppressor in a spontaneously immortalized but non-tumorigenic human epithelial cell line acts as a negative regulator of HPV transcription. Since continued growth of HPV transformed cell lines apparently depends on continued expression of HPV transforming genes (24), this model proposes that the formation of viable non-tumorigenic hybrids results from selection for cells in which this negative effect on HPV transcription has been abrogated. In support of this model, the HPV18 LCR cloned in front of a non-essential reporter gene was found to be down-regulated in non-tumorigenic hybrids.

The finding of YY1 site mutations in HPV16 associated cervical carcinomas, the evidence for a negative regulator of HPV transcription associated with tumor suppression, and our observation of elevated levels of YY1 in the non-tumorigenic hybrids all point to a role for YY1 in the pathway targeted by the tumor suppressor gene. It remains to define the mechanism of activation of this transcription factor in the non-tumorigenic cells. In principle, this activation could involve alterations in protein level, a post-translational modification such as phosphorylation, or a physical association with another protein. On Western blots, we have observed similar levels of immunoreactive YY1 protein in the tumorigenic and non-tumorigenic hybrids used in the present study. It therefore seems unlikely that the marked difference in YY1 DNA binding activity observed in the EMSA assays is a consequence of altered levels of YY1 protein. YY1 has been reported to associate with a variety of cellular proteins including Sp1 (25,26), c-myc (27), the nucleolar phosphoprotein B23 (28) and the adenovirus E1A associated protein p300 (29,30). Our results suggest the possibility that a protein targeting YY1 through a signal transduction pathway, and perhaps one of the YY1 associated proteins already identified, might represent the tumor suppressor gene product important in the HeLa–fibroblast hybrids.

## ACKNOWLEDGEMENTS

This work was supported by Public Health Service grant R01 CA-64114 from the National Cancer Institute. The early phases of this work were supported by NIH intramural funds to the Laboratory of Tumor Virus Biology, National Cancer Institute. J.L. gratefully acknowledges the support of Dr Peter M. Howley during the initial phases of this project.

## REFERENCES

1 Saxon, P.J., Srivatsan, E.S., and Stanbridge, E.J. (1986) *EMBO J.* **5**, 3461–3466.
2 Srivatsan, E.S., Benedict, W.F., and Stanbridge, E.J. (1986) *Cancer Res.* **46**, 6174–6179.
3 Dowdy, S.F., Lai, K.M., Weissman, B.E., Matsui, Y., Hogan, B.L.M., and Stanbridge, E.J. (1991) *Nucleic Acids Res.* **19**, 5763–5769.
4 zur Hausen, H. (1991) *Virology* **184**, 9–13.
5 Schlegel, R., Phelps, W.C., Zhang, Y.-L., and Barbosa, M. (1988) *EMBO J.* **7**, 3181–3187.
6 Bosch, F.X., Schwarz, E., Boukamp, P., Fusenig, N.E., Bartsch, D., and zur Hausen, H. (1990) *J. Virol.* **64**, 4743–4754.
7 Misra, B.C. and Srivatsan, E.S. (1989) *Am. J. Hum. Genet.* **45**, 565–577.
8 DiPaolo, J.A., Woodworth, C.D., Popescu, N.C., Notario, V., and Doninger, J. (1989) *Oncogene* **4**, 395–399.
9 Durst, M., Gallahan, D., Gilbert, J., and Rhim, J.S. (1989) *Virology* **173**, 767–771.
10 Rosl, F., Achtstatter, T., Bauknecht, T., Hutter, K., Futterman, G., and zur Hausen, H. (1991) *EMBO J.* **10,** 1337–1345.
11 Shrivastava, A. and Calame, K. (1994) *Nucleic Acids Res.* **22**, 5151–5155.
12 Bushmeyer, S., Park, K., and Atchison, M. (1995) *J. Biol. Chem.* **270**, 30213–30220.
13 Bauknecht, T., Jundt, F., Herr, I., Oehler, T., Delius, H., Shi, Y., Angel, P., and zur Hausen, H. (1995) *J.Virol.* **69**, 1–12.
14 May, M., Dong, X., Beyer-Finkler, E., Stubenrauch, F., Fuchs, P.G., and Pfister, H. (1994) *EMBO J.* **13**, 1460–1466.
15 Lichy, J.H., Modi, W.S., Seuanez, H.N., and Howley, P.M. (1992) *Cell Growth Diff.* **3**, 541–548.
16 Modi, W.S., Lichy, J.H., Dean, M., Seuanez, H.N., and Howley, P.M. (1991) *Cytogenet. Cell Genet.* **58**, 1968.
17 Richard, C.W., Boehnke, M., Berg, D.J., Lichy, J.H., Meeker, T.C., Hauser, E., Myers, R.M., and Cox, D.R. (1993) *Am. J. Hum. Genet.* **52**, 915–921.
18 Wheeler, T.T., Xiao, J.P., Dowdy, S.F., Stanbridge, E.J., and Young, D.A. (1991) *Oncogene* **6**, 1903–1907.
19 Frohman, M.A., RACE: Rapid amplification of cDNA ends. In: *PCR Protocols: A Guide to Methods and Applications*, M.A. Innis, Editor. 1990, Academic Press: San Diego. p. 28–38.
20 Gorman, C.M., Moffat, L.F., and Howard, B. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
21 Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T. (1978) *Cell* **15**, 1157–1174.
22 Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
23 Hyde-DeRuyscher, R., Jennings, E., and Shenk, T. (1995) *Nucleic Acids Res.* **23**, 4457–4465.
24 von Knebel Doeberitz, M., Bauknecht, T., Bartsch, D., and zur Hausen, H. (1991) *Proc. Natl. Acad. Sci. USA.* **88**, 1411–1415.
25 Seto, E., Lewis, B., and Shenk, T. (1993) *Nature* **365**, 462–464.
26 Lee, J.-S., Galvin, K.M., and Shi, Y. (1993) *Proc. Natl. Acad. Sci. USA.* **90**, 6145–6149.
27 Shrivastava, A., Saleque, S., Kalpana, G.V., Artandi, S., Goff, S.P., and Calame, K. (1993) *Science.* **262**, 1889–1892.
28 Inouye, C.J. and Seto, E. (1994) *J. Biol. Chem.* **269**, 6506–6510.
29 Shi, Y., Seto, E., Chang, L., and Shenk, T. (1991) *Cell* **67**, 377–388.
30 Lee, J., Galvin, K.M., See, R.H., Eckner, R., Livingston, D., Moran, E., and Shi, Y. (1995) *Genes Dev.* **9**, 1188–1198.