FEBS 22545 FEBS Letters 457 (1999) 429–436

# A novel downstream positive regulatory element mediating transcription of the human high mobility group (HMG) I-C gene

Kai-Yin Chau[a,1], Paola Arlotta[a,1], Umesh A. Patel[b], Colyn Crane-Robinson[b], Guidalberto Manfioletti[c], Santa Jeremy Ono[a,*]

[a] *The Schepens Eye Research Institute, Brigham and Women's Hospital, and Committee on Immunology, Harvard University, 20 Staniford St., Boston, MA 02114, USA*
[b] *Biophysics Laboratory, University of Portsmouth, Portsmouth P01 2DT, UK*
[c] *Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Università di Trieste, via Giorgieri 1, Trieste, Italy*

Received 28 July 1999

**Abstract The high mobility group (HMG) I proteins are small, non-histone chromosomal proteins that promote gene activation during development and within rapidly dividing cells. They do so by facilitating enhanceosome formation on inducible genes, via both protein/DNA and protein/protein interactions. The HMG I-C gene is tightly regulated, normally being expressed exclusively during embryonic development. However, HMG I-C expression is also observed frequently in a number of tumor types, and this expression has been shown to contribute to the malignant transformation process. With the aim of dissecting pathways that lead to aberrant expression of HMG I-C in tumor cells, we have analyzed HMG I-C gene regulation in the human hepatoma cell line PLC/PRF/5. One of the two HMG I-C transcripts detected in this cell line originates from a novel downstream initiation site at nucleotide −161 relative to the first methionine. Transcription from the downstream initiation site is mediated by a PRE located between nt −222 and −217. We show here that the Sp1 and Sp3 transcription factors interact with the PRE and transactivate the HMG I-C promoter in a cooperative fashion. This study provides the first characterization of this downstream HMG I-C promoter.**



*Key words:* Human high mobility group (HMG); Human high mobility group I-C gene (HMG I-C)

## 1. Introduction

High mobility group (HMG) I proteins are non-histone, chromatin-associated proteins that participate in nucleosomal assembly and the transcriptional regulation of RNA polymerase II genes [1–4]. The HMG I, Y and I-C proteins are referred to collectively as 'architectural' factors, as they have few or no direct transcription regulatory properties, but play critical roles in enhanceosome formation [5–7]. Unlike the similarly named, yet structurally distinct HMG proteins (containing an HMG box), the HMG I proteins interact with the minor groove of AT-rich DNA using three tandem DNA binding domains (also called AT hooks), located in the central portion of the polypeptides [8,9]. Via these interactions with DNA and with traditional transcription factors (e.g. Oct2, ATF-2, and NF-κB), the HMG I proteins have been shown to be critical for the generation of a higher order, multi-protein transcription complex on a variety of promoters (e.g. interferon-β, HLA-DRA and E-selectin) [10,11]. In addition, HMG I binding to DNA is also known to either bend DNA or reverse intrinsic bends, which has also been shown to promote the interaction of certain factors with their DNA recognition sequence [12].

An important role for HMG I proteins in transcriptional regulation was first shown with inducible genes whose gene products participate in the immune response [13]. One example involves the interferon-β gene, the expression of which is activated by viral infection [14]. In this case, HMG I(Y) proteins operate by reversing intrinsic bends within the promoter and by facilitating the interaction of ATF-2 and NF-κB with DNA and with each other [12,3]. A second example involves the human class II major histocompatibility complex (MHC) gene, HLA-DRA [15,16]. In this system, HMG I(Y) proteins interact with six discrete sites within the HLA-DRA proximal promoter. Interaction of the protein with three of these sites: the TATA box, the octamer element and AT-rich stretch between the consensus X and Y boxes, has been shown to be essential for induction of HLA-DRA gene expression by interferon-γ. In the case of the TATA box and the octamer element, HMG I(Y) contributes to gene activation by facilitating the loading of TBP or Oct-2A on these sites, respectively [15]. Interaction of HMG I(Y) with the 'interspace' element found between the X and Y boxes appears to contribute to induction of the gene by bending DNA and facilitating protein/protein interactions between X box and Y box binding factors. Finally, the HMG I-C protein appears to be important for the activation of the MuRantes promoter by viral infection, and has been shown to have very similar enhanceosome-promoting activities to HMG I(Y) [17].

The HMG I-C gene is expressed at high levels during embryonic development and is subsequently completely repressed in adult tissues [18,19]. Expression of the HMG I-C gene is reactivated in many tumors, and evidence now suggests that HMG I-C gene expression may contribute to cellular transformation [20]. Support for the hypothesis that HMG I-C gene expression is important for cellular growth comes from the finding that disruption of the HMG I-C gene results in a runted or 'pygmy' mouse with a specific blockade in adipogenesis [21]. In addition, the expression of antisense HMG I-C RNA effectively prevents neoplastic transformation of rat thyroid cells by retroviral oncogenes [22]. Most recently, this has been shown to operate through the induction of junB and fra-1, since inhibition of HMG I-C downregulates the expression of these factors, and since inhibition of these factors directly affects tumorigenicity [23]. Finally, the groups of Van de Ven

*Corresponding author. Fax: (1) (617) 912-0127.
E-mail: sjono@vision.eri.harvard.edu

[1] These authors contributed equally to this work.


PII: S0014-5793(99)01100-X

and Chada have made the remarkable finding that translocations frequently occur within the HMG I-C gene in a variety of benign mesenchymal tumors [24,25]. These translocations usually occur between exons 3 and 4 to either truncate the HMG I-C protein at the C-terminus of the third AT hook or to attach a heterologous coding sequence after the DNA binding domain to create a novel fusion protein. Thus, either the derepression of HMG I-C gene expression or the production of a truncated or fusion polypeptide may result in the activation of growth control genes and the dedifferentiation of terminally differentiated cells into rapidly growing cells which are destined to become benign or metastatic tumors [26].

With the aim of dissecting molecular mechanisms leading to HMG I-C gene expression in tumor cells, we have characterized the proximal promoter of the human gene in the context of the hepatoma cell line PLC/PRF/C [27]. This line has been previously shown to express high levels of HMG I-C transcripts. In this article, we report the identification of a novel downstream transcription initiation site and the characterization of its associated promoter. Mutagenesis of this downstream promoter indicates that an evolutionarily conserved repetitive sequence upstream of the start site acts as a major positive regulatory element, and binds the Sp1 and Sp3 transcription factors. Finally, transient transfection studies in SL-2 cells indicate that Sp1 and Sp3 can cooperate in the activation of this promoter.

## 2. Materials and methods

### 2.1. Sequence analysis

The entire nucleotide sequence of the human HMG I-C gene in the region discussed in this paper has been previously published by one of us [27]. Selected regions of the human HMG I-C gene 5′ untranslated region (UTR) were scanned for known *cis*-elements using the SIGSCAN v4.05 program at http://bimas.dcrt.nih.gov/molbio/signal [28].

### 2.2. Cell culture

Human hepatoma cells PLC/PRF/C and non-malignant human mammary epithelial cells HBL100 (from Marc E. Lippman, Lombardi Cancer Center, Georgetown University) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM GlutaMaxI and antibiotics. Schneider's *Drosophila melanogaster* cells SL-2 (purchased from ATCC) were cultured at 25°C in Schneider's Drosophila Medium supplemented with 10% fetal bovine serum and antibiotics. All reagents for cell culture were obtained from Gibco BRL (Gaithersburg, MD). HBL100 is a mammary line established in vitro from milk of an apparently healthy woman [29].

### 2.3. RNA analysis

Total RNA was prepared from exponentially growing cells using the acid phenol method [30]. For Northern blot analysis, RNA was electrophoresed in 1.5% formaldehyde-denaturing agarose gel, with the 0.24–9.5 kb RNA Ladder included in one lane as a size marker (Gibco BRL, Gaithersburg, MD). RNA was then transferred onto Biodyne B membrane (Pall, East Hills, NY) according to manufacturer's recommendations and fixed. The human HMG I-C gene-coding region was used as probe labeled with [$\alpha$-$^{32}$P]dCTP (ICN, Costa Mesa, CA) by random priming using the RediPrime DNA Labeling Kit (Amersham, Arlington Heights, IL) and purified by NucTrap Push Columns (Stratagene, La Jolla, CA). Hybridization was carried out in QuikHyb Hybridization Solution (Stratagene) at 68°C with 40 μg/ml sonicated salmon sperm DNA (Sigma, St. Louis, MO) and typically 10 ng/ml probe. The blot was stringently washed with 0.1×SSC, 0.1% SDS at 68°C for 15 min twice. Autoradiography was performed using Fuji Rx film with intensifying screen (Sigma, St. Louis, MO) at −70°C. The same blot was stripped after probing and then reprobed with the $^{32}$P-labeled 28S rRNA probe to control for equivalent RNA loading in each lane.

### 2.4. Primer extension analysis

The oligonucleotide: 5′-GGAGATGAGGTGATAGGGCTGGGG-3′ (spanning nt −75 to −98, Fig. 3C) (obtained from Genosys, Woodlands, TX) was end-labeled with [$\gamma$-$^{32}$P]ATP (ICN, Costa Mesa, CA) and T4 Polynucleotide Kinase (New England Biolabs, Beverly, MA) and then gel purified on a 13.3% polyacrylamide/1×TBE gel. 30 000 cpm of oligonucleotide was subsequently mixed with 20 μg of total RNA in 30 μl of hybridization solution (66% deionized formamide, 20 mM Tris-HCl, pH 7.4, 400 mM NaCl, 0.1% SDS and 1 mM EDTA) and heated at 85°C for 10 min. The solution was then incubated at 55°C for 8–12 h. The RNA and annealed primer were then ethanol precipitated, washed, vacuum dried and resuspended in 20 μl of reaction solution which contained 1×reverse transcriptase buffer, 0.5 mM of dNTPs, 20 units of RNasin ribonuclease inhibitor (Promega, Madison, WI) and 200 units of SuperScript II RNase H$^-$ Reverse Transcriptase (Gibco BRL, Gaithersburg, MD). The extension reaction was carried out at 42°C for 50 min and then stopped by adding EDTA to a final concentration of 25 mM. 10 μg of sonicated salmon sperm DNA and 2 μg of DNase-free RNase A (Sigma, St. Louis, MO) were added and incubated at 37°C for 30 min. The extension products were then purified with phenol/chloroform extraction, ethanol precipitated, washed, dried, resuspended in 2 μl of $dH_2O$ and 2 μl of sequencing gel loading dye. Samples were heated at 90°C for 2 min, then electrophoresed through 8% denaturing polyacrylamide (sequencing) gels. A sequencing ladder generated using the same oligonucleotide primer served as a marker. Autoradiography was performed to visualize the extension products.

### 2.5. Plasmid construction

Luciferase reporter plasmids were constructed by subcloning genomic fragments of the human HMG I-C gene 5′ UTR [31] into pGL2-Basic Luciferase Reporter Vector (Promega, Madison, WI). The remaining 5′ deletion constructs were prepared by nested deletion using the ExoIII/S1 deletion kit (MBI Fermentas, Amherst, NY). The orientation and nucleotide sequence of each construct were verified by DNA sequencing using Sequenase v.2 DNA Sequencing Kit (Amersham, Arlington Heights, IL). All restriction and modifying enzymes unless stated were obtained from New England Biolabs (Beverly, MA). Site-directed mutagenesis was conducted using the U.S.E. Mutagenesis Kit (Pharmacia Biotech, Piscataway, NJ) according to the manufacturer's protocol. An oligonucleotide carrying mismatched bases (underlined nucleotides): 5′-CCGCCACCGGCAGCGCCTGATCGATTCAAGGTAGTCCTCCCCTCTTCTCTTTTTGG-3′ was used to mutagenize nt −232 to −217 of the segment of the I-C gene promoter spanning from nt −289 to −31 of the luciferase reporter plasmid (Fig. 3C). Expression plasmids for Sp1 and Sp3 were kindly provided by Dr. Jonathan M. Horowitz at North Carolina State University.

### 2.6. Transfections, luciferase and β-galactosidase enzyme assays, and protein concentration

Plasmid DNA for transfection was prepared by Qiagen Plasmid kit (Chatsworth, CA). PLC/PRF/5 cells were seeded in 35 mm plates, grown for 20 h until they were 40% confluent. For each transfection, 2.5 μg of pGL2 (luciferase expression vector, Promega, Madison, WI) and 2.5 μg of each pCH110 Eukaryotic Assay Vector (served as an internal control to determine transfection efficiency by monitoring β-galactosidase expression, Pharmacia Biotech, Piscataway, NJ) were added to each transfection reaction using 15 μg of LipofectAMINE reagent according to the manufacturer's recommended procedures (Gibco BRL, Gaithersburg, MD). After 6 h cells were transferred to DMEM supplemented with 20% fetal calf serum and incubated for an additional 44 h at 30°C. Transfections were carried out in duplicate and repeated at least three times. Transfection of Drosophila SL-2 cells was performed using 6 μg of CellFectin reagent (Gibco BRL), and 1 μg of the luciferase reporter plasmid and 0.5 μg of the Sp1 or Sp3 expression plasmid. Cells were seeded in 35 mm plates to 40% confluence prior to transfection, and after 2 days cells were harvested.

For luciferase and β-galactosidase enzyme assays, the transfected cells were washed 2 times with PBS and then lysed in 150 μl of 1×reporter lysis buffer (Promega, Madison, WI). The cell lysates were harvested, frozen and thawed, and then spun. The supernatants were collected for use in luciferase and β-galactosidase assays. For luciferase analysis, 30 μl of the transfected cell extracts were mixed with 50 μl of room temperature equilibrated Luciferase Assay Reagent as indicated in protocols supplied by the manufacturer (Prom-

ega, Madison, WI). Luciferase activity was measured using a Monolight 2010 luminometer (Analytical Luminescence Laboratory, San Diego, CA). Light output was integrated over a 20 s period and displayed as relative light units (RLU). Background activity was found to be less than 0.1% of all readings. For β-galactosidase activity, the β-Galactosidase Enzyme Assay System was used following the instruction manual from the manufacturer (Promega, Madison, WI). All β-galactosidase activities were within the linear range of the absorbance curve. Protein concentration of cell lysates was assayed by BCA Protein Assay kit (Pierce, Rockford, IL).

*2.7. Preparation of nuclear extracts*

Nuclear extracts from PLC/PRF/C cells were prepared as follows: about $1\times10^6$ cells were harvested, washed twice with cold PBS and collected by centrifugation at 3500 rpm at 4°C for 5 min. The cells were resuspended in 400 μl of solution I (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA, 0.1 mM DTT, 0.5 mM PMSF) and lysed by passing them through a 25 gauge syringe. Nuclei were pelleted at 2000 rpm for 10 min, washed once with solution I and resuspended in 200 μl of solution II (solution I with 5% glycerol, 400 mM NaCl and without KCl). The suspension was rotated at 4°C for 30 min and then centrifuged at 14 000 rpm, 4°C, for 30 min. The resulting clear supernatant was stored at −80°C until use.

*2.8. Electrophoretic mobility shift assay (EMSA) and supershift assay*

A 20 bp long oligonucleotide 5′-CTCCTCCTCTCCTCCTCCTC-3′ (oligo A) extending from nt −234 to −215 on the 5′ UTR of HMG I-C gene, and its point-mutated version 5′-CTGATCGATTCAAGGTAGTC-3′ (m-oligo A) (Fig. 3C) were used in EMSA and supershift assay. Double stranded DNA fragments were end-labeled with [γ-$^{32}$P]ATP (ICN, Costa Mesa, CA) and T4 polynucleotide kinase (New England Biolabs, Beverly, MA). The probes were gel purified and 15 000 cpm were incubated for 25 min at room temperature with 4 μg of nuclear extract from PLC/PRF/C cell line, in the presence of 20 mM Tris-HCl pH 7.5, 75 mM KCl, 3.5 mM DTT, 20 nM $ZnCl_2$, 1 μg/μl BSA, 5% glycerol and 1 μg poly(dI-dC) (Pharmacia Biotech, Piscataway, NJ) in a final volume of 20 μl. For supershift experiments, 1 μg of antibody against Sp1, Sp3 or NF-κB p65 (Santa Cruz Biotechnology, Santa Cruz, CA) was incubated for 20 min at room temperature with the nuclear extract before adding the labeled probe. DNA/protein complexes were separated from free DNA on a 5% polyacrylamide gel in TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) at 4°C for 120 min at 100 V. After electrophoresis, gels were dried and autoradiographed.

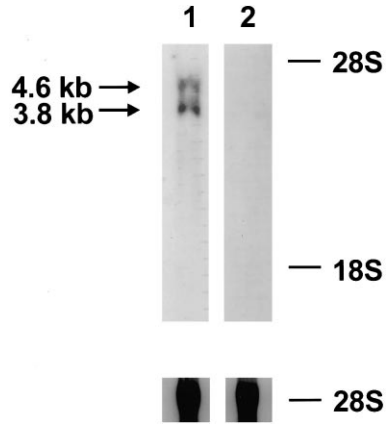


Fig. 1. Northern blot analysis of the hepatoma cells PLC/PRF/5, showing the presence of two HMG I-C transcript species. Total RNA (20 μg) isolated from human hepatoma cells PLC/PRF/5 (lane 1) and normal breast cell line HBL100 (lane 2) was resolved onto 1.5% denaturing gel and transferred to nylon membrane. The blot was hybridized with $^{32}$P-labeled human HMG I-C coding region (top panel) and then 28S rRNA probe (bottom panel). The sizes of the two I-C transcripts are indicated.

## 3. Results

*3.1. Two species of HMG I-C mRNA are detected in the PLC/PRF/5 hepatoma cell line*

The human HMG I-C cDNA was cloned previously from the PLC/PRF/5 hepatoma cell line [27], as it expresses extremely high levels of HMG I-C transcript. We have therefore chosen to dissect molecular mechanisms driving HMG I-C gene expression in this tumor cell line. Previous Northern blot analyses of RNAs isolated from this cell line, as well as from the human colorectal adenocarcinoma cell line DLD-1 and 15.5 dpc mouse embryos, detected a single transcript of approximately 4.1 kilobases [32]. These RNAs were resolved on low percentage (up to 1.2%) agarose gels. We have subsequently resolved a second, smaller transcript of 3.8 kb in

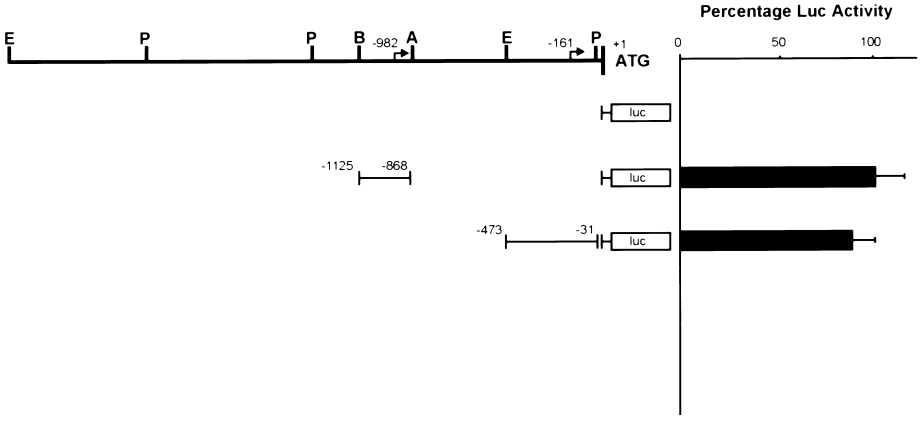


Fig. 2. Mapping of the HMG I-C gene promoters in the PLC/PRF/5 hepatoma cell line. Cells were transiently transfected using liposomes with constructs spanning different regions of the I-C gene 5′ UTR fused to luciferase reporter (diagram on the left). We number the nucleotide beginning immediately upstream from the first methionine ATG as +1, and the construct numbering parallels the position of the upstream region of I-C gene with respect to the upstream sequence from the translation start site. A promoter-less reporter plasmid was also transfected for negative control. Another SV40-driven β-galactosidase control plasmid was included in all transfection experiments. Luciferase activity was presented after normalization to β-galactosidase activity. The graph on the right shows the mean ± S.E.M. of at least three transfections with two independent batches of DNA.

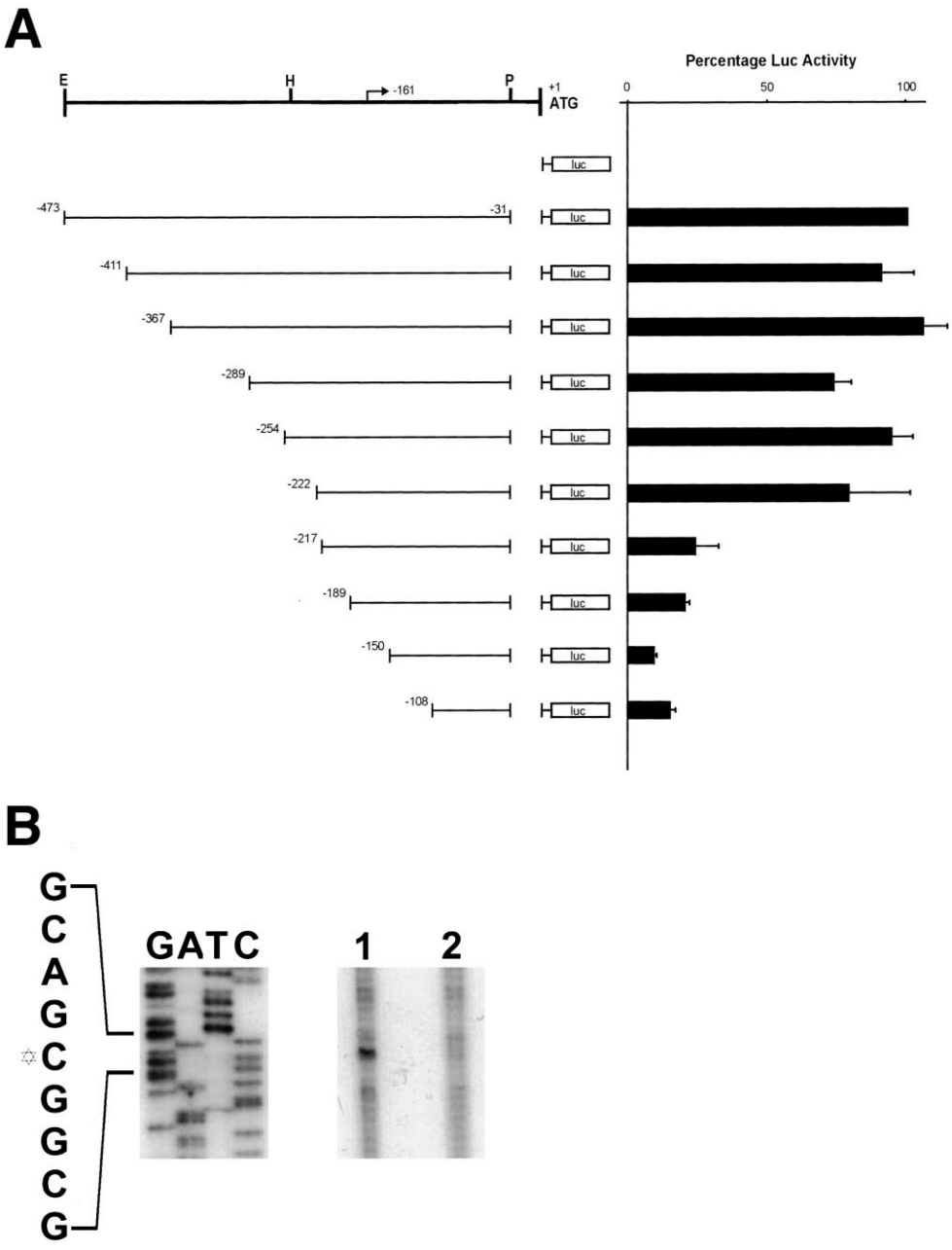


Fig. 3. Characterization of the downstream HMG I-C promoter (nt −473 to −31). A: Deletion analysis of the fragment by transfecting a 5′ deletion series of the fragment into PLC/PRF/5 cells. Data are a summary of at least three transfections from two independent plasmid preparation. The mean values and standard errors are shown. The PRE was demonstrated at nt −222 to −217, since deletion of this region reduced 3-fold the promoter activity of the preceding deletant. B: Primer extension analysis of the downstream promoter. Primer extension of 20 μg of total RNA prepared from PLC/PRF/5 (lane 1) and HBL100 (lane 2) was performed with $^{32}$P-end-labeled oligonucleotide complementary to the human I-C gene region nt −75 to −98 (relative to ATG) as described in the text. Primer extension products were analyzed onto a denaturing polyacrylamide gel, and size of the extended product were deduced from adjacent dideoxy-sequencing ladder. The Davidic star to the left of the sequence indicates the transcription start site (at nt −161). Experiments were repeated twice. C: Nucleotide sequence of the downstream HMG I-C gene promoter region. The nucleotide sequence is numbered in a way where +1 corresponds to the A of the initiator methionine codon, and residues preceding it are represented by negative numbers. The first nucleotide in each row is numbered on the left hand side, the number of the last nucleotide is indicated on the right hand side. Consensus transcription factor recognition sites are shown, among them Sp1/Sp3 binding sites are identified in the PRE (nt −222 to −217). Underlined is the sequence for gel-shift analysis, the bottom of which shown in gray represents the mutated sequence. The arrow depicts the location of the oligonucleotide annealed for primer extension. The bent arrow indicates the mapped transcriptional start site. Restriction sites are underlined. The methionine start codon is in upper case. D: EMSA and supershift analysis of region between nt −234 and nt −215 on the HMG I-C promoter. F, free probe; −, Sp1 and Sp3 show the result of band shift analysis in the absence or presence of antibody against Sp1 and Sp3 respectively. OLIGO A and m-OLIGO A are wild-type and point-mutated oligonucleotides as described in Section 2. The Sp1 supershifted complex is indicated by a double arrow.

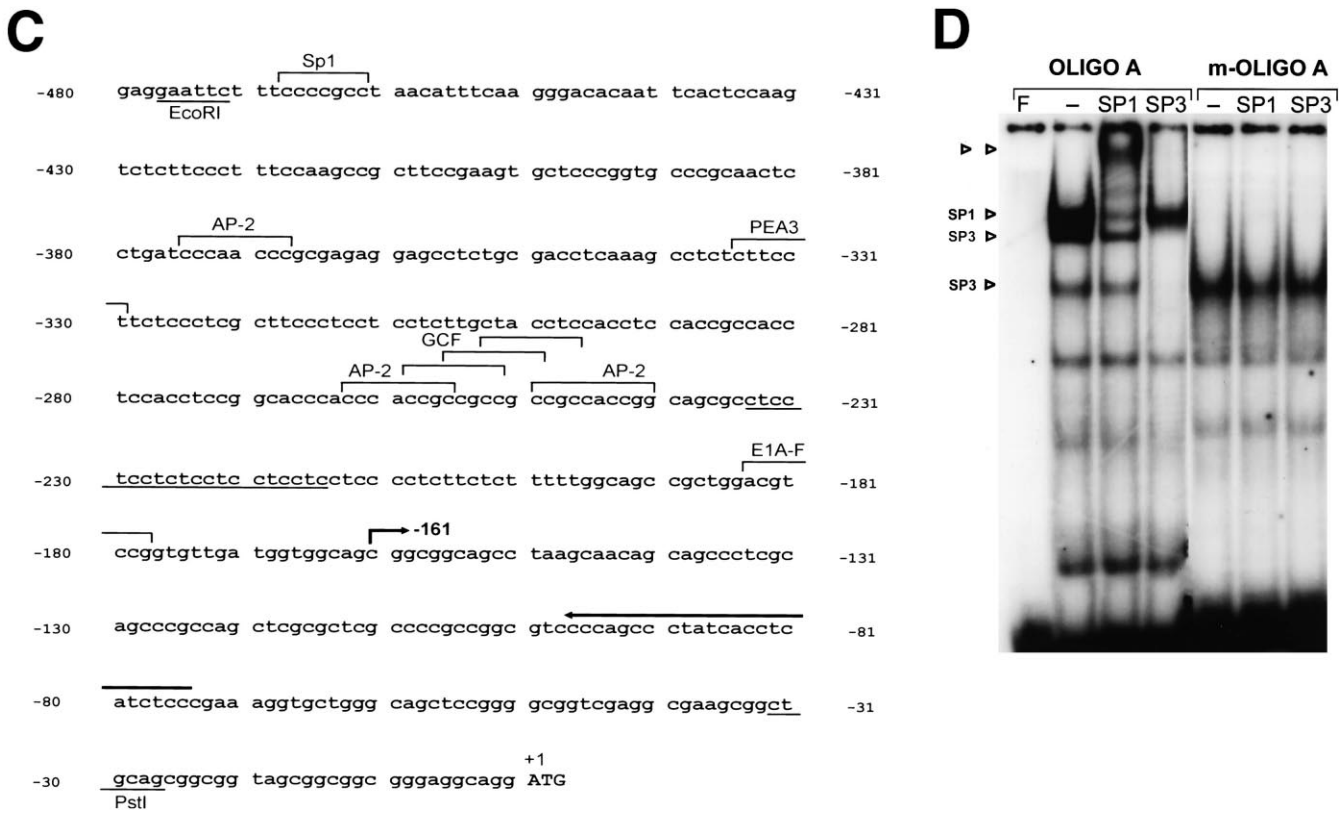


Fig. 3 (*continued*).

Northern blots of RNAs resolved on 1.5% denaturing agarose gels (Fig. 1). The smaller transcript is present at levels that are approximately 50% of the longer transcript. No HMG I-C transcripts were detected in the normal breast cell line HBL100. We therefore used this line as a negative control in subsequent experiments.

The existence of two forms of HMG I-C transcript in this cell line could be due to two distinct transcription initiation sites, two distinct transcription termination sites, and/or alternative splicing or processing of precursor RNAs. We ruled out the possibility that transcription was terminating at two sites, by performing multiple Northern blots using a series of DNA probes located downstream of the terminal HMG I-C exon (data not shown). We were unable to obtain selective hybridization to either the 4.6 or 3.8 kb band using any of the probes, strongly suggesting that termination occurs at either a single site, or in a restricted region. These data are in agreement with the data from the Chada and Van de Ven groups, and suggests that transcription terminates downstream of the second and third polyadenylation signals AAUAAA, producing a 3′ UTR of length about 3 kb [24,25]. In addition, it is highly unlikely that the two transcripts are due to alternative splicing within the coding region, as 0/10 cDNA clones isolated from a PLC/PRF/5 library, had internal deletions of any exon ([27], K.-Y. Chau, U.A. Patel and C. Crane-Robinson, unpublished observations).

### 3.2. *Identification of two upstream HMG I-C fragments with maximal promoter activity in the PLC/PRE/5 hepatoma cell line*

Several cDNAs were isolated from a library of PLC/PRF/5 hepatoma cell line and their 5′ termini mapped. The 5′ UTRs of all HMG I-C cDNAs cloned to date were less than 1 kb. Approximately half of the cloned cDNAs isolated from this cell line had a 5′ terminus approximately 800 bp upstream, and the other half cluster at less than 400 bp upstream of the first translational start codon. We therefore had a priori evidence that transcripts from this gene might initiate from two transcription initiation sites, and thus two promoters. We therefore generated a series of luciferase reporter constructs containing overlapping fragments within the proximal 3 kb upstream of the first ATG. These and several other reporter constructs were transfected into the PLC/PRF/5 hepatoma cell line, and promoter activities assessed (data not shown). Only two of the genomic fragments, one (spanning nt −1125 to −868) and the other (spanning nt −473 to −31) had strong promoter activity (Fig. 2). We therefore reasoned that both transcription initiation sites and positive regulatory elements driving HMG I-C gene expression in these cells would map within these fragments.

### 3.3. *Analysis of the downstream promoter fragment (nt −473 to −31) and localization of its transcription initiation site*

Since the downstream promoter fragment was unique to previous work in the literature (the upstream promoter has been previously described [32]), we focused our efforts on characterizing the downstream promoter. Unfortunately, there were no convenient restriction enzyme sites within the downstream fragment that could be used to generate the required progressive deletions at this location. Exonuclease III was therefore used to make progressive 5′ deletions, of which a fraction is shown in Fig. 3A. Truncations from nt −473 to

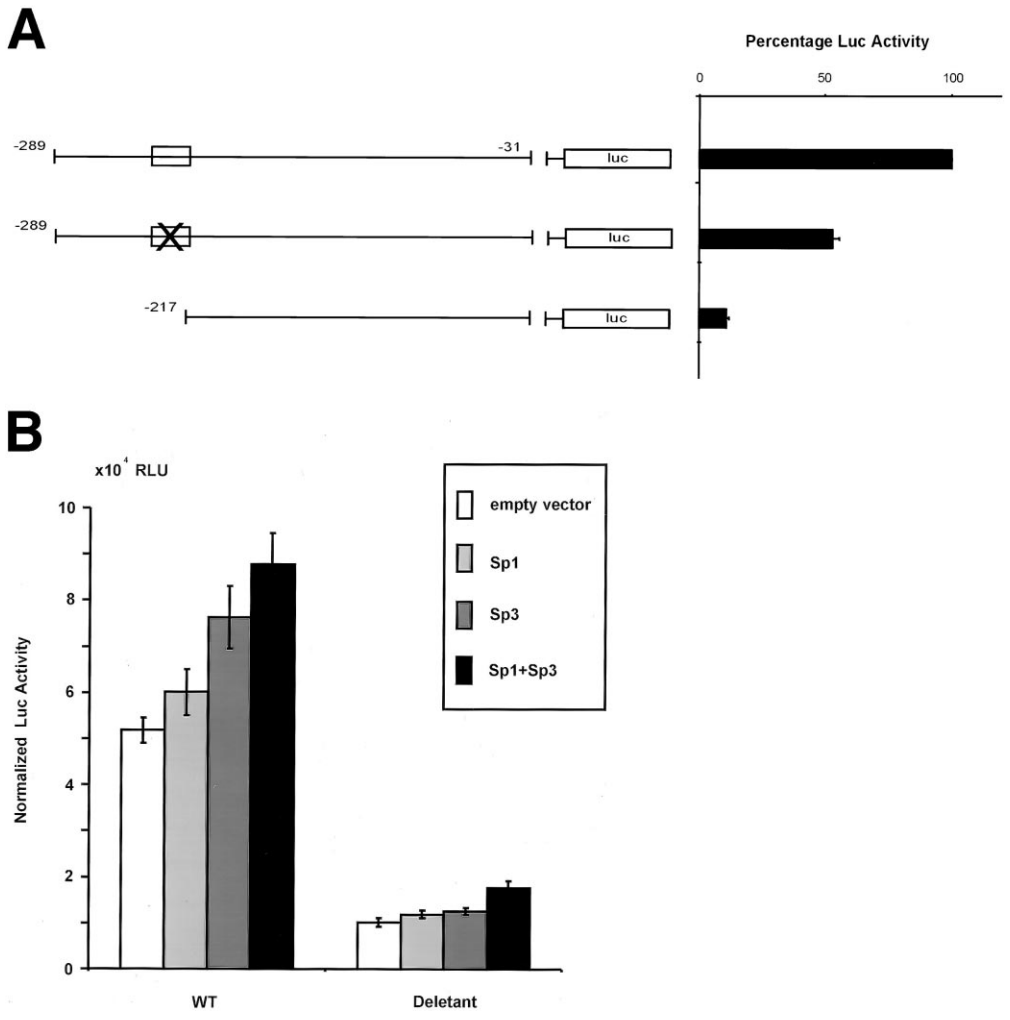


Fig. 4. Characterization of the Sp1/Sp3 site (CCTCC repeat) in the downstream HMG I-C promoter (nt −473 to −31). A: Transfection analysis of (top) the wild-type construct containing the I-C gene promoter region between nt −289 to −31 fused to the luciferase gene, (middle) site-directed mutant construct of region between nt −232 to −217 at the Sp1/Sp3 binding site, and (bottom) deletion mutant construct spanning region from nt −217 to −31, transfected into PLC/PRF/5 cells. The right panel represents the percentage luciferase activity of each reporter construct (as the mean with S.E.M. of at least three separate transfections from two independent plasmid preparation) relative to the wild-type construct. B: Transfection analysis of the wild-type (WT) and the deletion mutant (deletant) constructs into *Drosophila* SL-2 cells, co-transfected with Sp1 or Sp3 expression plasmids, shown as normalized luciferase activity. Data are the mean of at least three separate transfections from two independent plasmid preparations.

−222 had little effect on promoter activities within the PLC/PRF/5 cell line. However, deletion to nt −217, and all further deletions resulted in a dramatic decrease in promoter activity. We interpreted these data to indicate that either a critical positive regulatory element (PRE), or a transcription initiation site is located near nt −222 and −217. We therefore mapped the region surrounding these nucleotides for transcription initiation site(s). Although several spurious extension products were observed in RNAs extracted from both the PLC/PRF/5 and HBL100 cell lines (Fig. 3B, lanes 1 and 2), we reproducibly observed one unique extension product in RNAs from the PLC/PRF/5 cell line corresponding to a downstream transcription initiation site (Fig. 3B, lane 1). The sequence of this region and the position of the start site are shown in Fig. 3C. The location of the primer binding site is also shown as a bold antiparallel arrow. As was the case with the upstream start, no TATA boxes are found near this site.

### 3.4. *Sp1 and Sp3 DNA binding proteins interact with the downstream HMG I-C PRE in a sequence-specific manner*

We inspected the nucleotide sequences within the downstream PRE for known *cis*-elements. Analysis of the downstream PRE flanking sequence between nt −473 and −161 revealed a series of potential Sp1, AP-2, GCF and other binding sites (Fig. 3C). Since deletion of the promoter sequences to nt −222 had little effect on promoter activity in the PLC/PRF/5 cell line, we feel that the potential *cis*-elements upstream of this site are irrelevant for PRE activity in this particular cell line. However, the potential Sp1 sites located downstream of nucleotide −222 were of immediate interest. We therefore performed EMSA using an oligonucleotide spanning nt −234 to nt −215 (oligo A) to determine whether Sp1 and/or Sp3 factors were present in the PLC/PRF/5 extracts, and interact with this sequence. The experiment revealed four specific DNA/protein complexes (Fig. 3D). Supershift assays using antibodies against Sp1 and Sp3 showed that

both Sp1 and Sp3 are indeed part of the shifted DNA/protein complexes (Fig. 3D). No supershifted bands were observed when an unrelated antibody (against NF-κB p65) was used (data not shown). Similar experiments were repeated using a point-mutated oligonucleotide (m-oligo A) carrying mutations that alter both Sp1 and Sp3 consensus binding sites. As shown in the second half of Fig. 3D, no supershifted complexes were observed when using the mutated oligonucleotide.

### 3.5. *Definition of the downstream HMG I-C PRE and its transactivation by Sp1 and Sp3*

Since the downstream PRE serves as a binding site for two known transcriptional activating proteins (Sp1 and Sp3), we tested the ability of reporter constructs which either lack this site, or harbor a mutated version of the Sp1/Sp3 binding site, to drive luciferase gene expression in PLC/PRF/5 cells (Fig. 4A). Deletion of the downstream PRE (to −217) severely abrogates promoter activity (10% of wild-type activity), and mutagenesis of the PRE (with the mutant sequence being the same as that used in m-oligo A in Fig. 3D) results in a 50% inhibition of promoter activity (Fig. 4A). Taken together, these data indicate that the downstream PRE is critical for transcription from the downstream initiation site, and that the Sp1/Sp3 sequence (likely in cooperation with unknown sequences downstream of −289, but upstream of −222) is an important component of that promoter.

Since we had shown that the Sp1/Sp3 sequences are important for transcription from the downstream PRE, and that Sp1 and Sp3 proteins are components of nucleoprotein complexes forming on these oligonucleotides, we next tested whether the Sp1 and Sp3 proteins could transactivate this promoter in Sp1/Sp3-deficient SL-2 cells. Fig. 4B shows that while Sp1 and Sp3 can only weakly transactivate the wild-type reporter construct that the two factors can cooperate to result in maximal transcription. A reporter lacking the downstream PRE is unresponsive to Sp1, Sp3, or Sp1 and Sp3.

## 4. Discussion

Cancers frequently arise from the inappropriate activation of growth control genes either from translocation events or from derepression of gene expression. Classic examples of translocation-mediated gene activation/transformation are the chromosome 8:14 translocation in B-cell lymphoma (such that the immunoglobulin heavy chain enhancer drives c-myc gene expression), and translocation into band 3q27 (BCL-6) in diffuse large-cell lymphoma [33,34]. The direct role of inappropriate expression of oncogenes in cellular transformation is well documented, especially in several transgenic mouse models of cancer [35–38]. Aberrant expression of either full length or C-terminally truncated HMG I-C mRNA is a frequent occurrence in neoplastic tissue [19,39–42]. This is particularly true for benign tumors of mesenchymal origin, but also has been observed in a variety of other benign and malignant tumors [43]. Two biological phenomena suggest that this aberrant expression might contribute to the transformed phenotype. First, deletion of the HMG I-C gene results in a *pygmy* or runt phenotype, indicating that the HMG I-C gene product is required for normal growth and development [21]. Second, inhibition of HMG I-C gene expression via expression of antisense HMG I-C mRNA arrests the in vitro growth of thyroid tumor cell lines, and interferes with their tumorigenicity in nude mice [22]. We have therefore begun to probe molecular mechanisms driving HMG I-C gene expression in the hepatoma cell line PLC/PRF/5 with the aim of understanding how this gene is reactivated in tumor tissue.

This study establishes that HMG I-C gene transcription initiates from two major discrete sites within this hepatoma cell line. The downstream initiation site reported in this study is unique from that previously reported, and the promoter elements driving this activity have therefore been characterized [33]. A single major transcription initiation site was mapped by Chada's group within the DLD-1 cell line [32]. This start is 118 bp 3′ of the upstream start site mapped our study (data not shown). Several minor extension products were also reported observed in the DLD-1 line using two primers. Thus, our current mapping of two major transcription initiation sites in using PLC/PRF/5, along with Chada's previous mapping of multiple unrelated initiation sites in DLD-1, extends the heterogeneity in initiation sites at this upstream promoter. This is consistent with what is observed in many TATAA-less promoters. The distinct initiation sites in the two cell lines is interesting as it opens up the possibility of cell-type specific activation mechanisms.

While previous work has mapped transcription initiation sites in the DLD-1 cell line, this work represents the first analysis of cis-elements regulating transcription from any start within the HMG I-C gene. The data presented here demonstrate that transcription from each start site is controlled by distinct positive regulatory elements. The upstream PRE is located within nt −1125 to −1065 (data not shown), and the downstream PRE within nt −222 to −217. Since our upstream start is near Chada's start, we have elected to focus our attention in this study on the novel downstream start site.

Sequence analysis of the downstream PRE indicates the presence of multiple known *cis*-elements. Most notably, binding sites for AP-2, Sp1/Sp3 and GCF are clustered near the functional PRE (Fig. 3C). Gel mobility shift experiments indicate that both Sp1 and Sp3 (but not AP-2 – data not shown) proteins interact with the downstream PRE (Fig. 3D). Interestingly, distinct complexes in the EMSAs contained Sp1 and Sp3, suggesting that distinct combinations of these factors with other uncharacterized factors might mediate transcription from this promoter. The existence of other nearby factors (probably binding just upstream of the PRE) is supported by the significant activity of the reporter construct harboring a site-specific mutation in the PRE (relative to the wild-type −289 reporter), the dramatic affect of deletion to −217, and the relatively weak transactivation of the reporters by Sp1 and Sp3.

The data presented here suggest that activation of HMG I-C expression in this hepatoma cell line is mediated by two major PREs, which are distinct from those mapped in human colorectal adenocarcinoma [32]. Thus, different molecular pathways are likely to activate HMG I-C gene expression in different tumor types. In support of this, we have recently found that retinoblastomas initiate transcription from two additional unique sites (K.-Y. Chau and S.J. Ono, unpublished observations). Careful studies in various tumor cell types and at different developmental stages will therefore be required to properly understand HMG I-C gene regulation.

In the future we will focus on the biochemical purification of additional DNA binding proteins interacting with the two PREs, and their molecular cloning. The proteins interacting

with the upstream PRE are completely unknown. The CCGCCC consensus identified associated with both the upstream and downstream PREs are present in promoters of several genes involving growth control like EGF receptor, SV40 early, c-myc ([34] and references therein, Fig. 7) and Ha-ras 1 [49]. The downstream PRE contains the characterized Sp1/Sp3 binding site with homology to the S1_HS sites (5′-TCCTCCTCC-3′) found in the EGF receptor [34]. Therefore, these are likely to be significant with respect to growth control, and are of particular interest. It is known that Sp1 can interact with the sequence homologous to hypersensitive sites and stimulate the in vitro transcription of templates containing reiterated copies of this sequence [34,35], and it is therefore significant that we have shown that both Sp1 and Sp3 can interact with this site in PLC/PRF/5 cells.

It is also noteworthy that HMG I-C gene induction has recently been shown to be dependent on the RAF-1/p44/p42 mitogen-activated protein kinase-dependent pathway in salivary epithelial cells [44]. Additional future experiments will investigate whether oncogenic Raf-1 can alter DNA/protein interaction at either PRE identified in this study.

*Acknowledgements:* We thank other members of our lab: Drs. A.M. Keane-Myers, C.S. Parry, G. Zhou, M. Radosevich and K.-W. Cheung-Chau for advice. We also thank Drs. C.M. Read and I. Manfield, Biophysics Labs, University of Portsmouth, UK; Dr. D.K.-L. Lee, Hong Kong Polytechnic University; Dr. P.H.-Y. Lam, Hong Kong University of Science and Technology for invaluable discussions at the beginning of this work. This work was supported by National Institute of General Medical Science Grant RO1 GM49661, the Lucille P. Markey Charitable Trust, the Research to Prevent Blindness, the 'Associazione Italiana per la Ricerca sul Cancro', and internal grants from the Hong Kong University of Science and Technology and the Hong Kong Polytechnic University. G. Manfioletti was supported by 'Fondazione Italiana per la Ricerca sul Cancro' when he was a Visiting Scientist of the Ono Lab. S.J. Ono was a JSPS Visiting Professor at Kyoto University, during the preparation of the manuscript. P. Arlotta is a fellow of the Telethon Fund, Italy, and K.-Y. Chau a Fellow of the Research to Prevent Blindness, America Fund (PD97054).

## References

[1] Johns, E.W. (1982) The HMG Chromosomal Proteins, Academic Press, New York.
[2] Bustin, M., Lehn, D.A. and Landsman, D. (1990) Biochim. Biophys. Acta 1049, 231–243.
[3] Huth, J.R., Bewley, C.A., Missen, M.S., Evans, J.N.S., Reeves, R., Gronenborn, A.M. and Clore, G.M. (1997) Nature Struct. Biol. 4, 657–665.
[4] Farnet, C.M. and Bushman, F.D. (1997) Cell 88, 483–492.
[5] Wolffe, A.P. (1994) Science 264, 1100–1101.
[6] Grosschedl, R., Giese, K. and Pagel, J. (1994) Trends Genet. 10, 94–100.
[7] Mantovani, F., Covaceuszach, S., Rustighi, A., Sgarra, R., Heath, C., Goodwin, G.H. and Manfioletti, G. (1998) Nucleic Acids Res. 26, 1433–1439.
[8] Johnson, K.R., Lehn, D.A., Elton, T.S., Barr, P.J. and Reeves, R. (1988) J. Biol. Chem. 263, 18338–18342.
[9] Yie, J., Liang, S., Merika, M. and Thanos, D. (1997) Mol. Cell. Biol. 17, 3649–3662.
[10] Thanos, D. and Maniatis, T. (1995) Cell 83, 1091–1100.
[11] Carey, M. (1998) Cell 92, 5–8.
[12] Falvo, J.V., Thanos, D. and Maniatis, T. (1995) Cell 83, 1101–1111.
[13] Bustin, M. and Reeves, R. (1996) Prog. Nucleic Acid Res. 54, 35–100.
[14] Thanos, D. and Maniatis, T. (1992) Cell 71, 777–789.
[15] Abdulkadir, S.A., Krishna, S., Thanos, D., Maniatis, T., Strominger, J.L. and Ono, S.J. (1995) J. Exp. Med. 182, 487–500.
[16] Abdulkadir, S.A. and Ono, S.J. (1995) FASEB J. 9, 1429–1435.
[17] Lokuta, M.A., Maher, J., Noe, K.H., Pitha Shin, M.L. and Shin, H.S. (1996) J. Biol. Chem. 271, 13731–13738.
[18] Zhou, X., Benson, K.F., Przybysz, K., Liu, J., Hou, Y., Cherath, L. and Chada, K. (1996) Nucleic Acids Res. 24, 4071–4077.
[19] Rogalla, P., Drechsler, K., Frey, G., Hennig, Y., Helmke, B., Bonk, U. and Bullerdiek, J. (1996) Am. J. Pathol. 149, 775–779.
[20] Wunderlich, V. and Bottger, M. (1997) J. Cancer Res. Clin. Oncol. 123, 133–140.
[21] Zhou, X., Benson, K.F., Ashar, H.R. and Chada, K. (1995) Nature 376, 771–774.
[22] Berlingieri, M.T., Manfioletti, G., Santoro, M., Bandiera, A., Visconti, R., Giancotti, V. and Fusco, A. (1995) Mol. Cell. Biol. 15, 1545–1553.
[23] Vallone, D., Battista, S., Pierantoni, G.M., Fedele, M., Casalino, L., Santoro, M., Viglietto, G., Fusco, A. and Verde, P. (1997) EMBO J. 16, 5310–5321.
[24] Schoenmakers, E.F.P.M., Wanschura, S., Mols, R., Bullerdiek, J., Van den Berghe, H. and Van de Ven, W.J.M. (1995) Nature Genet. 10, 436–444.
[25] Ashar, H.R., Schoenberg Fejzo, M., Tkachenko, A., Zhou, X., Fletcher, J.A., Weremowicz, S., Morton, C.C. and Chada, K. (1995) Cell 82, 57–65.
[26] Tkachenko, A., Ashar, H.R., Meloni, A.M., Sandberg, A.A. and Chada, K. (1997) Cancer Res. 57, 2276–2280.
[27] Patel, U.A., Bandiera, A., Manfioletti, G., Giancotti, V., Chau, K.-Y. and Crane-Robinson, C. (1994) Biochem. Biophys. Res. Commun. 201, 63–70.
[28] Prestridge, D.S. (1991) CABIOS 7, 203–206.
[29] Vanhamme, L. and Szpirer, C. (1988) Carcinogenesis 9, 653–655.
[30] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[31] Chau, K.-Y., Patel, U.A., Lee, K.-L.D., Lam, H.-Y.P. and Crane-Robinson, C. (1995) Nucleic Acids Res. 23, 4262–4266.
[32] Ashar, H.R., Cherath, L., Przybysz, K.M. and Chada, K. (1996) Genomics 31, 207–214.
[33] Dalla-Favera, R., Martinotti, S., Gallo, R.C., Erikson, J. and Croce, C.M. (1983) Science 219, 963–967.
[34] Ye, B.H., Lista, F., Lo Coco, F., Knowles, D.M., Offit, K., Chaganti, R.S. and Dalla-Favera, R. (1993) Science 262, 747–750.
[35] Hinrichs, S.H., Nerenberg, M., Reymolds, R.K., Khoury, G. and Jay, G. (1987) Science 237, 1340–1343.
[36] Koike, K., Hinrichs, S.H., Isselbacher, K.J. and Jay, G. (1989) Proc. Natl. Acad. Sci. USA 86, 5615–5619.
[37] Kim, C.M., Koike, K., Saito, I., Miyamura, T. and Jay, G. (1991) Nature 351, 317–320.
[38] Green, J.E., Baird, A.M., Hinrichs, S.H., Klintworth, G.K. and Jay, G. (1992) Am. J. Pathol. 140, 1401–1410.
[39] Giancotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A. and Goodwin, G.H. (1989) Exp. Cell Res. 184, 538–545.
[40] Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B. and Goodwin, G.H. (1991) Eur. J. Biochem. 198, 211–216.
[41] Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautiere, P., Cary, P., Crane-Robinson, C., Coles, B. and Goodwin, G.H. (1991) Nucleic Acids Res. 19, 6793–6797.
[42] Giancotti, V., Bandiera, A., Ciani, L., Santoro, D., Crane-Robinson, C., Goodwin, G.H., Boiocchi, M., Dolcetti, R. and Casetta, B. (1993) Eur. J. Biochem. 213, 825–832.
[43] Kottickal, L.V., Sarada, S., Ashar, H., Chada, K. and Nagarajan, L. (1998) Biochem. Biophys. Res. Commun. 242, 452–456.
[44] Li, D., Lin, H.H., McMahon, M., Ma, H. and Ann, D.K. (1997) J. Biol. Chem. 272, 25062–25070.