Functional analysis of the promoter region of the human phosphotyrosine phosphatase activator gene: Yin Yang 1 is essential for core promoter activity

Veerle JANSSENS, Christine VAN HOOF, Ivo DE BAERE, Wilfried MERLEVEDE and Jozef GORIS[1]

Afdeling Biochemie, Faculteit Geneeskunde, Katholieke Universiteit Leuven, Herestraat 49, B-3000 Leuven, Belgium

The phosphotyrosine phosphatase activator (PTPA) has been isolated as an *in vitro* regulator of protein phosphatase 2A. Human PTPA is encoded by a single gene, the structure and chromosomal localization of which have been determined in our previous work. Here we describe the further isolation, sequencing and functional characterization of the *PTPA* promoter region. In agreement with its ubiquitous expression, the *PTPA* promoter displays several characteristics of housekeeping genes: it lacks both a TATA-box and a CAAT-box, it is very GC-rich and it contains an unmethylated CpG island surrounding the transcription initiation site. Transient transfection experiments in different cell types with several truncated chimaeric luciferase reporter gene plasmids revealed the importance of the region between positions −67 and −39 for basal promoter activity. This region coincides remarkably well with the determined CpG island. Further analysis of this region demonstrated the presence of a Yin Yang 1 (YY1) binding motif at positions −52 to −44. Binding of YY1 to this sequence is demonstrated in bandshift and DNase I footprinting experiments. Another YY1 binding motif is found in the 5′ untranslated region, at positions +27 to +35. Mutations in either of these sites, abolishing YY1 binding *in vitro*, have differential effects on promoter activity. Point mutations in both sites completely abolish promoter activity. Moreover, induction of promoter activity by co-transfection with a YY1 expression plasmid is fully dependent upon the presence of both intact YY1 binding sites. Thus YY1 apparently mediates basal transcription of the human PTPA gene through two binding sites within its proximal promoter.

Key words: CpG island, protein phosphatase 2A, TATA-less promoter, transcriptional regulation.

INTRODUCTION

Protein phosphatase 2A (PP2A) is a major serine/threonine phosphatase implicated in cell-cycle progression, responses to signalling, cell transformation and cell differentiation [1,2]. Apart from its apparent phosphoserine/threonine phosphatase activity, PP2A also exhibits low but detectable phosphotyrosine phosphatase (PTPase) activity [3,4]. There is evidence that this PTPase activity may be physiologically relevant. For example, it was observed that the dimeric form of PP2A ($PP2A_D$) complexed with polyoma middle T/small t displays elevated PTPase activity [5], suggesting that the intrinsic PTPase activity of $PP2A_D$ is stabilized by middle T/small t. Moreover, PP2A can autodephosphorylate and re-activate itself after phosphorylation on Tyr^{307} [6–8]. *In vitro* the PTPase activity of $PP2A_D$ can be upregulated specifically by a protein called PTPA (PTPase activator) [9–11]. PTPA can modulate $PP2A_D$ activity only in the presence of ATP/Mg^{2+}. The exact mechanism of activation is not clear, although phosphorylation can be ruled out [9,12]. PTPA has been cloned from rabbit and human cDNA libraries [12], and homologues have been found in *Xenopus*, *Drosophila*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* [13]. An alignment of these different PTPA clones reveals highly conserved regions of amino acids, which are essential for PTPA activity [13]. Its ubiquitous and high expression in diverse mammalian tissues and cell lines [14] and its apparent high conservation throughout evolution point to an important function for this protein.

So far, no regulation of PTPA itself has been described. In order to investigate possible regulation at the transcriptional level, Van Hoof et al. [15] cloned the human PTPA gene, determined its transcriptional start site and started the analysis of part of its 5′ flanking sequence to investigate its potential as a promoter. We have now further isolated and sequenced about 3 kb of this 5′ upstream region and analysed the whole promoter region more in detail. We found that, in agreement with its ubiquitous expression, the PTPA gene promoter displays several features characteristic of housekeeping genes. The functionality of the promoter was analysed in transfection experiments with chimaeric luciferase reporter gene plasmids in three different cell types. By deletion mapping, a sequence essential for core promoter activity was identified. Using mutational analysis, co-transfection experiments and *in vitro* DNA binding assays, we demonstrated that this sequence contains a functional binding site for transcription factor Yin Yang 1 (YY1). Interestingly, another YY1 binding motif was found in the 5′ untranslated

Abbreviations used: CMV, cytomegalovirus; EMSA, electrophoretic mobility shift assay; GST, glutathione S-transferase; NFI/CTF, nuclear factor I/CCAAT binding transcription factor; PP2A, protein phosphatase type 2A; $PP2A_D$, dimeric form of PP2A; PTPase, phosphotyrosine phosphatase; PTPA, PTPase activator; Sp1, specificity protein 1; UTR, untranslated region; YY1, Yin Yang 1.

[1] To whom correspondence should be addressed (e-mail jozef.goris@med.kuleuven.ac.be)

The complete *PTPA* promoter sequence (nt −2681 to +207) has been deposited in the Genbank/EMBL/DDBJ Nucleotide Sequence Databases under accession number AF134185.

region (UTR), which also seems to contribute, although to a lesser extent, to *in vivo* promoter activity. Thus we show that YY1 acts as a positive regulator of PTPA gene expression.

EXPERIMENTAL

Isolation and sequencing of *PTPA* upstream sequences

The genomic clones HA and HF, both containing the *PTPA* upstream sequences [15], were analysed by sequential restriction digestions and Southern blotting. The restriction fragments generated were mapped by hybridization with oligonucleotide probes (5′-end-labelled with T4 polynucleotide kinase and [γ-^{32}P]ATP) that were complementary to previously mapped DNA fragments of known composition. The fragments of interest were cloned into pBluescript II SK+ (Stratagene) or pGEM-7Zf(+) (Promega) vectors and sequenced by the dideoxy-chain-termination method [16] using the T7 DNA Polymerase Sequencing kit (Pharmacia) with either T3, T7 and SP6 primers or specific internal primers.

Preparation of genomic DNA and Southern blotting for methylation analysis

High-quality genomic DNA was isolated from liver HepG2 cells as described by Majumdar et al. [17]. A 30 μg sample of DNA was double-digested with *Sac*I/*Nae*I, *Sac*I/*Nar*I or *Sac*I/*Eag*I, or with *Sac*I alone as a control. Electrophoresis was conducted on 0.8% (w/v) agarose gels, and transfer to a Zeta-Probe Membrane (Bio-Rad) was carried out by vacuum blotting. The membrane was hybridized with a T4 polynucleotide kinase ^{32}P-labelled oligonucleotide corresponding to nt -1870 to -1827 (5′-TTTGAATGAGCAGGTCCAGGGGACTCTGGGCAAGTCCCTCCT TG-3′) at 60 °C for 24 h. Washing was performed as described by Sambrook et al. [18].

Construction of reporter plasmids

A 1 kb *Apa*I fragment containing 648 bp of the 5′ flanking sequence, the complete first exon and part of the first intron was subcloned into pBluescript II SK+. Digestion with *Pst*I/*Xho*I and treatment with exonuclease III (Erase-a-Base kit; Promega) generated a 3′ deletion and removed the *PTPA* start codon. The resulting fragment, comprising bases -648 to $+38$, was cloned into pGEM-7Zf(+) as a *Kpn*I/*Sac*I fragment. Treatment of this construct with *Sac*I, the Klenow fragment of *Escherichia coli* DNA polymerase I (Klenow polymerase) and *Xho*I generated a fragment that could be cloned into the pGL2-Basic vector (Gene Light Plasmid; Promega) treated with *Hin*dIII, Klenow polymerase and *Xho*I. The resulting luciferase reporter construct was named $-648/+38$-pLuc. Digestion of $-648/+38$-pLuc with *Sac*I/*Xho*I, treatment with exonuclease III (Erase-a-Base) and subsequent re-ligation generated a set of 5′-deleted reporter constructs, comprising $-394/+38$ bp and $-102/+38$ bp respectively. Digestion of this construct with *Sph*I and treatment with Klenow polymerase and *Hin*dIII generated a small fragment that could be inserted in the *Sma*I/*Hin*dIII-treated pGL2-Basic vector. This construct was named $-67/+38$-pLuc. The shortest reporter construct ($-39/+38$-pLuc) was made by digesting $-648/+38$-pLuc with *Sma*I and autoligating the remaining plasmid. Two longer reporter constructs were made by subcloning the 1.8 kb *Sac*I/*Apa*I and the 0.8 kb *Nhe*I/*Apa*I genomic fragments into $-648/+38$-pLuc, resulting in a $-2356/+38$-pLuc and a $-1565/+38$-pLuc construct respectively. All constructs generated were confirmed by sequencing.

Cell culture

All cell lines were purchased from the American Type Culture Collection (Rockville, MD, U.S.A.). HepG2 cells (human liver hepatoma) and L929 cells (murine lung fibroblasts) were cultured at 37 °C and 5% CO_2 in Dulbecco's modified Eagle's

Table 1 Oligonucleotides used in bandshift and footprint experiments

Nucleotides in bold indicate the point mutations introduced.

Name	Ref.	Sequence
YY1-CON	[20]	5'-CGCTCCGCGGCCATCTTGGCGGCTGGT-3'
		3'-GGCGCCGGTAGAACCGCCGACCA-5'
NFI-CON	[21]	5'-ATTTTGGCTACAAGCCAATATGAT-3'
		3'-AACCGATGTTCGGTTATACTA-5'
PTPA-YY1(u)		5'-CCCGCACTGA**C**AT**G**GCCGTCGCCCGGGTCCGC-3'
		3'-GTGACT**G**TA**C**CGGCAGCGGGCCCAGGCGCGCA-5'
PTPA-YY1(u)MUT		5'-CCCGCACTGA**A**AT**T**GCCGTCGCCCGGGTCCGC-3'
		3'-GTGACT**T**TA**A**CGGCAGCGGGCCCAGGCGCGCA-5'
PTPA-YY1(d)		5'-CGTTAATAGGCTTGCTCCCTGAGCGCCCCGCACCGA**C**AT**G**GCGGCCGTC-3'
		3'-TTATCCGAACGAGGGACTCGCGGGGCGTGGCT**G**TA**C**CGCCGGCAG-5'
PTPA-YY1(d)MUT		5'-CGTTAATAGGCTTGCTCCCTGAGCGCCCCGCACCGA**A**AT**T**GCGGCCGTC-3'
		3'-TTATCCGAACGAGGGACTCGCGGGGCGTGGCT**T**TA**A**CGCCGGCAG-5'

Table 2 Sequences of oligonucleotides used to generate YY1 binding site mutations in the context of the −102/+38 minimal *PTPA* promoter

Combinations of oligonucleotides 1, 2, 3a, 4, 5, 6a, 7, 8, 9a and 10a generated –102/+38 YY1(u)mut; combinations of 1, 2, 3b, 4, 5, 6b, 7, 8, 9b and 10b generated –102/+38 YY1(d)mut; and combinations of 1, 2, 3a, 4, 5, 6a, 7, 8, 9b and 10b generated –102/+38 YY1(u,d)mut. The 3′ overhang of oligonucleotide 2 and the 5′ overhang of oligonucleotides 10a and 10b were used for cloning the different fragments in the *Nhe*I/*Kpn*I sites of pGL2-Basic.

	Sequence
1	5-'CCTAGCGCTTGGCGGCCGTTGGCGCGCA-3'
2	3'-CATGGGATCGCGAACCGCCGGCAACC-5'
3a	5'-TGCGCACAGCGCGCCCCGCACTGAAATT-3'
3b	5'-TGCGCACAGCGCGCCCCGCACTGACATG-3'
4	3'-GCGCGTACGCGTGTCGCGCGGGGCGTGA-5'
5	5'-GCCGTCGCCCGGGTCCGCGCGTCCGCCG-3'
6a	3'-CTTTAACGGCAGCGGGCCCAGGCGCGCA-5'
6b	3'-CTGTACCGGCAGCGGGCCCAGGCGCGCA-5'
7	5'-CGCGCCGGCCGTTAATAGGCTTGCTCCC-3'
8	3'-GGCGGCGCGCGGCCGGCAATTATCCGAA-5'
9a	5'-TGAGCGCCCCGCACCGACATGGCGGCCG-3'
9b	5'-TGAGCGCCCCGCACCGAAATTGCGGCCG-3'
10a	3'-CGAGGGACTCGCGGGGCGTGGCTGTACCGCCGGCGATC-5'
10b	3'-CGAGGGACTCGCGGGGCGTGGCTTTAACGCCGGCGATC-5'

medium (Gibco/BRL), supplemented with 10% (v/v) fetal calf serum (Gibco/BRL), 2 mM L-glutamine (Gibco/BRL), 100 units/ml penicillin and 100 mg/ml streptomycin (Gibco/BRL). Saos-2 cells (human osteosarcoma) were cultured in McCoy's 5A medium (Gibco/BRL) supplemented with 15% (v/v) fetal calf serum, 100 units/ml penicillin and 100 mg/ml streptomycin, at 37 °C and 5% CO_2.

Transient transfection assays

Transient transfections were performed in six- or 24-well dishes using Fugene reagent (Boehringer) or calcium phosphate/DNA co-precipitation as described by Sambrook et al. [18]. The pGL2-Promoter plasmid (Promega), containing the simian virus 40 promoter, and the promoterless pGL2-Basic vector were used as positive and negative controls respectively in separate but parallel assays. To evaluate transfection efficiencies, cells were co-transfected with a reporter vector containing the β-galactosidase gene driven by the cytomegalovirus (CMV) promoter. The amount of co-transfected DNA was one-tenth of the total DNA used in the transfection. At 48 h after transfection, cells were harvested in 250 or 100 μl of passive lysis buffer (Promega). Protein concentrations were measured using the bicinochoninic acid system (Pierce). After concentration adjustment, luciferase activity was measured with the Luciferase Assay System (Promega) according to the manufacturer's protocol using a MicroLumat LB96P (Berthold) with automated injection. Reagents for measurement of β-galactosidase activity were obtained from Tropix. For each construct at least two different DNA preparations (QIAGEN) were tested, and each transfection was repeated at least three times.

Electrophoretic mobility shift assays (EMSAs)

Nuclear extracts of cultured cells (L929 and HepG2) were prepared as described by Dignam et al. [19]. Bacterial expression, purification and thrombin cleavage of YY1–GST fusion proteins (where GST is glutathione S-transferase) in the pGEX-system (Pharmacia) were performed according to the manufacturer's instructions. The protein was found to be > 80% pure on Coomassie Blue-stained SDS/PAGE gels.

Complementary synthetic DNA oligonucleotides (Table 1) were denatured at 95 °C for 5 min, left to cool down slowly to 4 °C for annealing and labelled by a fill-in reaction with Klenow polymerase in the presence of [α-^{32}P]dATP and/or [α-^{32}P]dCTP. Reaction mixtures containing 40000 c.p.m. of labelled probe were pre-incubated on ice for 10 min in binding buffer [10 mM Tris/HCl, pH 8.0, 1 mM dithiothreitol, 12% glycerol, 0.5 mM EDTA, 80 mM KCl, 0.05% Triton X-100, 0.2 mg/ml BSA, 50 ng/μl poly(dI/dC)]. For competition experiments, at this point the appropriate unlabelled double-stranded DNA competitor oligonucleotides were included in the reaction mixture. Then 1–5 μg of nuclear extract or 10–100 ng of purified recombinant protein was added and the incubation was continued for 30 min at room temperature. For supershift analysis, 10 μl of anti-YY1 antibody or 10 μl of preimmune serum (TransCruz Supershift Reagent; Santa Cruz Biotechnology) was added 15 min after protein addition, and incubation was continued overnight at 4 °C. Samples were analysed on non-denaturing 5% (w/v) polyacrylamide gels in 0.5 × TBE (1 × TBE = 45 mM Tris/borate/1 mM EDTA)/0.05% Triton. After electrophoresis, gels were dried and bands were visualized by autoradiography.

In vitro DNase I footprinting

The DNA fragment used was a *Hin*dIII/*Pst*I restriction fragment from a pBluescript II SK + construct, containing the −60/+50 fragment of the minimal *PTPA* promoter, cloned in the *Eco*RV restriction site in the sense direction. The −60/+50 fragment was generated by 'hCG' PCR [22] with PWO polymerase (Boehringer), using two specific primers (5′-CCCGCACTGAC-ATGGCCGTCGCCCGGGTCCGC-3′ and 5′-CACAGCGAA-GACGGCCGCCATGTCGG-3′) and the HF genomic clone as template. The use of this particular PCR technique was necessary because of the high GC content of the sequence. End-labelling of the fragment (in this case in the sense strand) was done by a Klenow polymerase fill-in reaction of the *Hin*dIII site in the presence of [α-^{32}P]dATP and [α-^{32}P]dCTP. For the footprinting reactions, 50000–100000 c.p.m. of radiolabelled probe was incubated for 20 min at 20 °C in 50 μl of binding mixture containing 8.5% glycerol, 50 mM NaCl, 2.5 mM $MgCl_2$, 10 mM Hepes, 0.05 mM EDTA, 2% polyvinyl alcohol, 20 ng/μl poly(dI/dC), 1 mM dithiothreitol and different amounts of recombinant thrombin-cleaved YY1–GST. Different dilutions of DNase I (Boehringer) in 10 mM $CaCl_2$ and 12.5 mM $MgCl_2$ were added and the incubation was continued for exactly 1 min. The reaction was stopped by adding 100 μl of DNase I stop solution (1% SDS, 200 mM NaCl and 20 mM EDTA). When competition assays were performed, 5–10 pmol of competitor DNA was added in the binding mixture and the DNase I incubation was prolonged slightly (for 10–15 s). After phenol/chloroform ex-

traction and ethanol precipitation, the samples were washed, dried and subjected to electrophoresis on 8 % (w/v) denaturing (7 M urea) polyacrylamide gels in 0.5 × TBE. Gels were fixed in 10 % (v/v) acetic acid, washed, dried and exposed. As sequence reference, a guanine-specific chemical cleavage reaction was performed on the same probe as described by Maxam and Gilbert [23].

Site-directed mutagenesis

To introduce point mutations into the YY1 binding sites in the context of the −102/+38-pLuc plasmid, standard PCR techniques were not suitable because of the high GC content and sequence repetitiveness of the promoter region. Therefore another approach was chosen. Five pairs of complementary and partially overlapping synthetic DNA oligonucleotides, some containing the desired mutations (Table 2), were phosphorylated by T4 polynucleotide kinase (Pharmacia), denatured, annealed and ligated. The resulting fragments were then ligated into pGL2-Basic digested with *Nhe*I/*Kpn*I and named −102/+38 YY1(u)mut-pLuc, −102/+38 YY1(d)mut-pLuc and −102/+38 YY1(u,d)mut-pLuc [in these plasmids the upstream (u), the downstream (d) or both YY1 sites are mutated]. Sequencing of the constructs confirmed the introduction of the different mutations.

RESULTS

Further isolation and sequence analysis of the 5′ flanking region of the human PTPA gene

We analysed the genomic clones HA and HF, known to contain the upstream sequences of the PTPA gene [15], by Southern blotting, as described in the Experimental section. Fragments of these clones, containing the known extreme 5′ end of the PTPA gene, were subsequently cloned and sequenced. Analysis of the complete DNA sequence from nt −2681 to +207 (Genbank accession no. AF134185) revealed the presence of several potential binding sites for transcription factors. No consensus CAAT or TATA boxes are present in the correct position relative to +1 and, unlike many other TATA-less promoters, no pyrimidine-rich initiator binding site [5′ (C/T)(C/T)A^{+1}N(A/T)(C/T)(C/T) 3′] is present either [24]. Potentially important transcription factor binding sites include some that play a role in cell proliferation and differentiation, such as those for E2A (E-box family), Ets-1, ELP (embryonal long terminal repeat binding protein), MAF (mammary activating factor) and PEA3 (polyoma enhancer factor 3) (Pu-box family), glucocorticoid response element (GRE) and peroxisome-proliferator-activated receptor (PPAR) (nuclear receptor family), YY1, Myb, CCAAT enhancer binding protein (C/EBP), p53 and activator protein-1 (AP-1). However, the majority of the transcription factor binding sites detected are very GC-rich [including potential Sp1 (specificity protein 1)/Sp3, Krox-24/EGR-1 (early growth response gene product 1), GCF (GC factor) and AP-2 (activator protein-2) sites], reflecting the high GC content of the promoter, another well known characteristic of many TATA-less genes [25].

Identification of an unmethylated CpG island in the proximal promoter region

The apparent high GC content of the promoter prompted us to evaluate the sequence for the presence of CpG islands. CpG islands contain unusually high contents of CpG dinucleotides, which occur very rarely in vertebrate DNA. They are commonly defined as DNA regions of at least 200 bp in length with an average GC content above 50 %, and a ratio of observed to expected CpG doublets of greater than 0.6 [26]. They coincide frequently with the 5′ ends of housekeeping genes and of some tissue-specific genes, and are postulated to constitute sites of interaction between transcription factors and promoters [27]. Normally, they remain unmethylated, and their aberrant methylation has been shown to promote transcriptional inactivation [28].

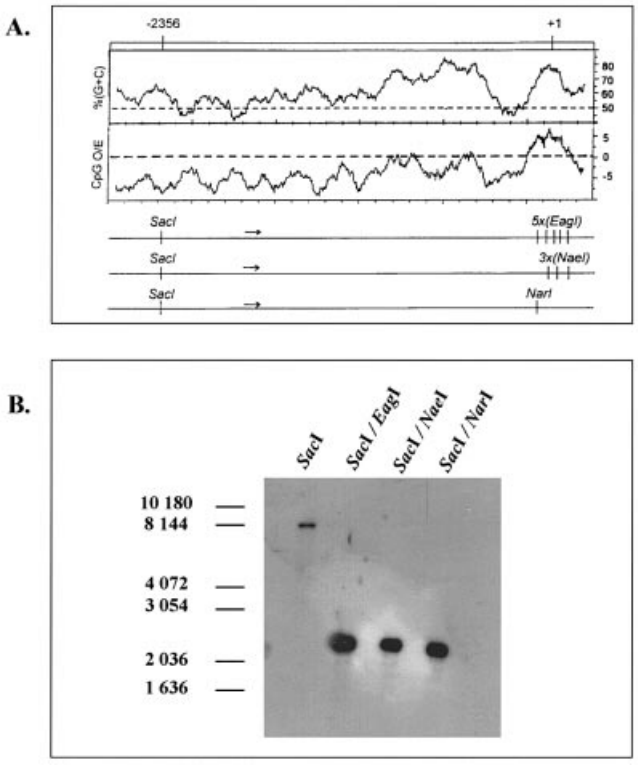

Figure 1 CpG island characterization

(**A**) The upper plot shows the percentage G + C content of the 5′ flanking sequence of the human PTPA gene. The lower plot shows the observed/expected frequency of CpG doublets per 100 bp. This ratio was calculated as: (number of CpG × *N*)/(number of C × number of G), where *N* is the number of nucleotides in the sequence being analysed. In both plots the depicted values were calculated using a 100 bp window ($N = 100$) moving along the sequence at 3 bp intervals. The broken lines indicate the minimal criteria required for a CpG island (respectively 50% GC content and observed/expected CpG frequency of $\geqslant 0.6$). Below the plots, the expected restriction fragments using three methylation-sensitive enzymes and *Sac*I as a 5′ limiting enzyme are indicated for each enzyme. The arrow indicates the position of the probe used for hybridization of the Southern blot. (**B**) Southern blot of genomic DNA isolated from liver HepG2 cells probed after digestion with *Sac*I (lane 1) and double-digestion with *Sac*I/*Eag*I (lane 2), *Sac*I/*Nae*I (lane 3) and *Sac*I/*Nar*I (lane 4). The sizes of DNA standards (kb) are indicated on the left.

Statistical analysis of the *PTPA* promoter region revealed a putative CpG island with an average GC content of 71 % and an observed/expected CpG frequency of up to 7 between nt −158 and +91 (Figure 1A). Its methylation state was checked by digesting genomic DNA, extracted from PTPA-expressing HepG2 cells, with three methylation-sensitive restriction enzymes (*Nar*I, *Nae*I and *Eag*I), and with *Sac*I as a 5′ limiting enzyme. Since *Nar*I, *Nae*I and *Eag*I sites are only present within the CpG island and not in the rest of the promoter region, we may expect to detect a fragment of approx. 2.2 kb with a promoter-specific probe if the sequence is not methylated. However, if the sequence is methylated, we should see a larger fragment (max. 8 kb), the size of which will depend upon the presence of *Nar*I, *Nae*I and

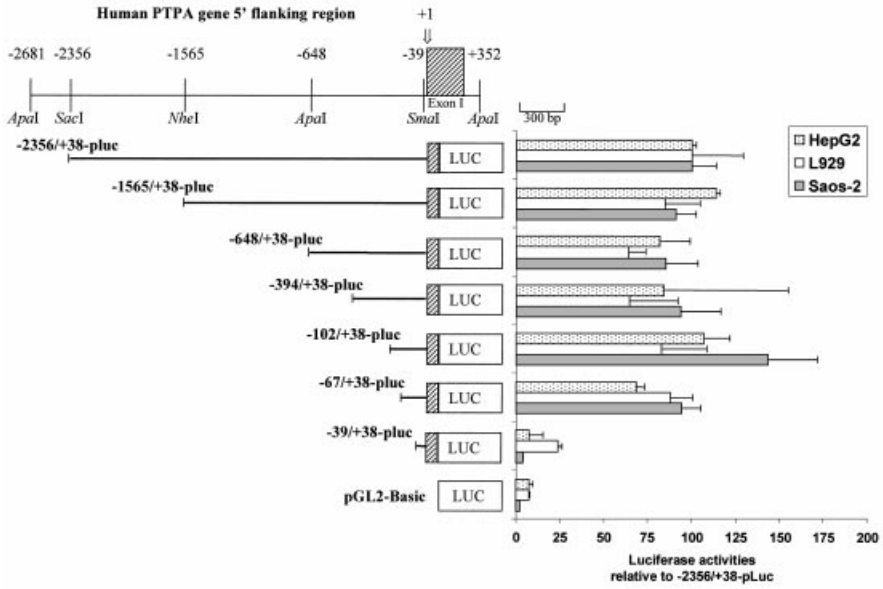

Figure 2 Activities of *PTPA* promoter fragments cloned in the pGL2-Basic vector in transient transfections in HepG2, L929 and Saos-2 cells

The promoter fragments have their 3′ ends at +38 and their 5′ ends at respectively −2356, −1565, −648, −394, −102, −67 and −39 nt relative to the transcription initiation site (a schematic representation of the deletion constructs is shown on the left). The measured values were first corrected for endogenous luciferase activity by subtraction of the blank (luciferase activity of untransfected cells) and subsequently for differences in transfection efficiency by adjusting them to the average β-galactosidase value of all the samples in the same transfection experiment. The reported values are relative corrected activities of the promoter constructs compared with the corrected activity of −2356/+38-pLuc, which was set at 100%. For each value, the S.E.M. of at least three independent experiments is indicated.

*Eag*I restriction sites downstream of the CpG island. The results of the Southern blot clearly indicate that the sequence is not methylated, since a fragment of approx. 2.2 kb was detected with all three methylation-sensitive enzymes (Figure 1B). Consequently, in HepG2 cells the PTPA gene contains an active unmethylated CpG island in the proximal promoter region (nt −158/+91).

Functional delineation of the minimal human PTPA gene promoter

To localize the region essential for transcriptional activity, we generated a set of chimaeric constructs containing different genomic fragments of the human PTPA gene fused to a luciferase reporter gene. These constructs were transfected into human liver HepG2 cells, murine lung L929 cells and human bone Saos-2 cells, along with a CMV/β-galactosidase expression vector as an internal control for varying transfection efficiencies within one cell line. Luciferase and β-galactosidase activities were measured 48 h later. In none of these three PTPA-expressing cell lines could significant differences be observed between the luciferase activities of −2356/+38-pLuc, −1565/+38-pLuc, −648/+38-pLuc, −394/+38-pLuc, −102/+38-pLuc and −67/+38-pLuc (Figure 2). In L929 cells, overall promoter activities seemed to be somewhat lower than in Saos-2 and HepG2 cells, but these differences can probably be explained in terms of different transfection efficiencies among different cell lines. However, the severe decrease in luciferase activity of the −39/+38-pLuc plasmid compared with that in the longer reporter constructs (Figure 2) was consistent in all cell lines, indicating that the region between nt −67 and nt −39 is crucial for basal expression of the human PTPA gene. Interestingly, this region coincides remarkably well with the determined CpG island.

YY1 binds to the promoter region essential for promoter activity

Between nt −67 and −39, the region essential for promoter activity (see above), two well-conserved consensus transcription factor binding sites are present: a perfect (9/9) YY1 binding site at positions −52/−44 and an imperfect (11/14) nuclear factor I/CCAAT binding transcription factor (NFI/CTF) binding site at positions −49/−36. To assess whether these sites constitute real transcription factor binding sites, bandshift assays (EMSAs) were performed with a synthetic oligonucleotide comprising nt −60/−25 of the *PTPA* promoter [PTPA-YY1(u); Table 1] and nuclear extracts of L929 cells.

First, it was shown that these extracts contain functional NFI/CTF and YY1 (Figure 3A). Incubation of oligonucleotides containing known NFI/CTF and YY1 binding sites (NFI-CON and YY1-CON respectively; Table 1) with these extracts resulted in two clear retarded complexes of different electrophoretic mobility. Therefore it is clear that both transcription factors are present in our L929 extracts and that both can bind to their respective control probes under the experimental conditions used.

Upon incubation of the PTPA-YY1(u) probe with these extracts under the same conditions, a clear retarded band could be observed (Figure 3B, lane 4), which could be competed by an excess of unlabelled PTPA-YY1(u) (lanes 7 and 8) and by an excess of unlabelled YY1-CON (lanes 5 and 6), but not by an excess of NFI-CON (results not shown). Furthermore, incubation of radiolabelled YY1-CON with the same extracts resulted in a retarded complex (lane 2) that obviously co-migrated with the observed PTPA-YY1(u) binding complex (lane 4), whereas the complex formed on NFI-CON migrated significantly more slowly (Figure 3A).

In order to confirm whether YY1 indeed interacts with its candidate YY1 binding site at positions −52/−44, EMSAs were

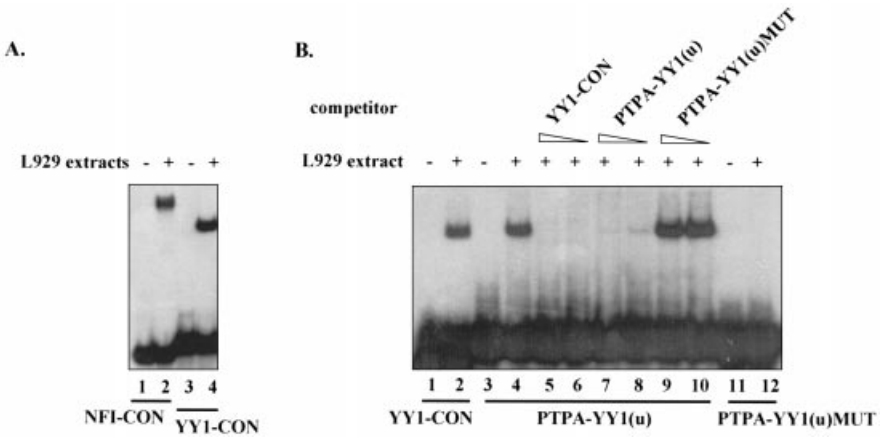

Figure 3 Identification of a YY1 binding site in the core promoter in bandshift assays with L929 nuclear extracts

(**A**) Labelled oligonucleotides are indicated below the lanes: NFI-CON (lanes 1 and 2) and YY1-CON (lanes 3 and 4) contain known binding sites for NFI/CTF and YY1 respectively (see also Table 1). In lanes 2 and 4, 2 μg of L929 nuclear extract was added. Lanes 1 and 3 are controls without protein addition. (**B**) Labelled oligonucleotides are indicated below the lanes: YY1-CON (lanes 1 and 2), PTPA-YY1(u), containing nucleotides $-60/-25$ of the *PTPA* promoter (lanes 3–10), and PTPA-YY1(u)MUT, identical to PTPA-YY1(u) except for two point mutations in the core sequence of the putative YY1 binding site at nt $-52/-44$ (lanes 11 and 12). Lanes 1, 3 and 11 are controls without L929 nuclear extracts. In lanes 2, 4 and 12 no competitor DNA was added. In lanes 5–10 a 200-fold or 100-fold molar excess of unlabelled competitor DNA was added (indicated above the gel): the consensus YY1 oligonucleotide (lanes 5 and 6), PTPA-YY1(u) itself (lanes 7 and 8) or the mutated oligonucleotide PTPA-YY1(u)MUT (lanes 9 and 10).

performed with PTPA-YY1(u)MUT. This oligonucleotide is identical to PTPA-YY1(u), except for two point mutations in the putative YY1 binding site, which destroy its 5′-CATN-3′ binding core [29] in both strands (G^{-52}**A**C**AT**G**G**CC^{-44} was mutated to G^{-52}**AA**A**TT**GCC^{-44}) (Table 1). In contrast with wild-type PTPA-YY1(u), no retarded complex could be observed with PTPA-YY1(u)MUT (Figure 3B, lane 12). In addition, an excess of unlabelled PTPA-YY1(u)MUT failed to compete with labelled PTPA-YY1(u) for the PTPA-YY1(u) binding protein (Figure 3B, lanes 9 and 10). Taken together, these data strongly support the presence of a YY1 binding site in the minimal *PTPA* promoter at positions $-52/-44$.

Identification of a second YY1 binding site in the 5′ UTR

Examination of the complete proximal promoter sequence between nt -67 and $+38$ revealed the presence of a second putative YY1 binding motif at nt $+27/+35$, again perfectly matching the YY1 consensus sequence. As for the upstream YY1 binding site, we assessed whether YY1 could bind to this site in bandshift assays. For this purpose, a synthetic oligonucleotide, comprising nt $-8/+41$ of the *PTPA* promoter [PTPA-YY1(d); Table 1], was labelled and incubated with L929 nuclear extracts. A clear retarded band could be observed that disappeared upon addition of unlabelled YY1-CON and unlabelled PTPA-YY1(d) (results not shown). Following the same approach as for the identification of the upstream YY1 site, we next introduced similar point mutations in the downstream YY1 site: G^{+27}A**C**-A**TG**GCG^{+35} was mutated to G^{+27}**AA**A**TT**GCG^{+35} (Table 1). The resulting mutated oligonucleotide, PTPA-YY1(d)MUT, not only failed to bind YY1, but also failed to compete with labelled PTPA-YY1(d) for protein binding (results not shown).

To confirm the former data, bandshifts were performed with a *PTPA*-promoter-specific probe, comprising nt $-60/+50$ and thus containing both putative YY1 binding sites. In the presence of low L929 protein concentrations, binding complexes were formed in which only one YY1 site was occupied (one retarded band was seen), whereas in the presence of high protein concentrations the second site became occupied by YY1 as well (two retarded bands were seen). In both cases the specificity of protein binding was confirmed in competition assays with unlabelled YY1-CON (results not shown).

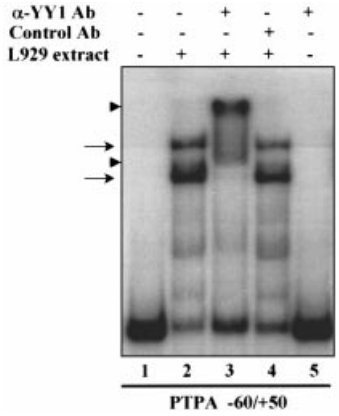

Figure 4 Supershifts of *PTPA* promoter binding complexes with YY1-specific antibodies

A probe containing nt $-60/+50$ of the *PTPA* promoter was incubated without (lane 1) or with (lanes 2–4) 10 μg of L929 nuclear extract. In lane 3, 10 μl of a polyclonal antibody specific for YY1 (α-YY1 Ab) was added. Lane 4 is a control to which the same amount of preimmune serum was added. In lane 5 the probe was incubated with 10 μl of anti-YY1 antibody without L929 extracts. The retarded complexes in the absence of antibody are indicated by arrows, whereas the arrowheads indicate the complexes formed in the presence of antibody.

Supershift assays with YY1-specific antibodies

To further confirm the identity of the binding protein as being YY1, we performed supershift experiments with the $-60/+50$ *PTPA*-promoter-specific probe, L929 nuclear extracts and a polyclonal antibody specific for YY1 (Figure 4). The two retarded complexes obtained by addition of relatively high amounts of

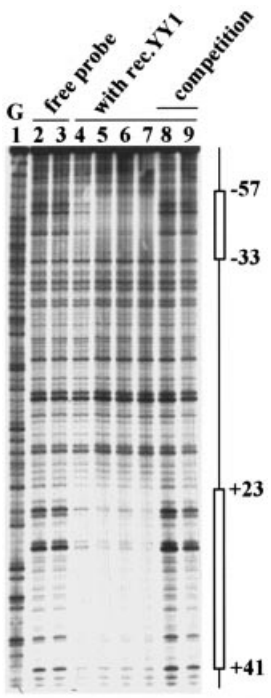

Figure 5 ***In vitro*** **DNase I footprinting of the minimal** ***PTPA*** **promoter**

A *Hin*dIII/*Pst*I restriction fragment, comprising nucleotides −60/+50 of the *PTPA* promoter, was end-labelled in the sense strand. It was treated for exactly 1 min with 1×10^{-2} or 5×10^{-3} units of DNase I (lanes 2 and 3), or with 0.1 unit of DNase I when recombinant (rec.) YY1 was added (lanes 4–9). In lanes 8 and 9, 10 pmol of unlabelled YY1-CON competitor DNA was added and the DNase I incubation was prolonged by 10 s. Lane 1 represents the G-specific Maxam–Gilbert sequencing reaction performed on the same probe. Boxes represent the regions protected from DNase I digestion, and their positions within the sequence are indicated.

L929 nuclear proteins (lane 2) were clearly supershifted upon addition of anti-YY1 antibody (lane 3), whereas addition of control antibody (preimmune serum; lane 4) did not alter the mobilities of the retarded complexes. Addition of antibody alone did not result in any retardation of the probe (lane 5).

In vitro DNase I footprinting of the minimal *PTPA* promoter

To substantiate the former results, *in vitro* DNase I footprinting experiments were performed. As protein source in this assay we used a recombinant version of human YY1, whose *in vitro* binding capacity was first tested in a bandshift assay with YY1-CON, PTPA-YY1(u) and PTPA-YY1(d). Addition of the protein to these probes resulted in the formation of retarded complexes indistinguishable from the complexes obtained by using L929 nuclear proteins (results not shown).

When this recombinant YY1 was used in footprinting experiments with the −60/+50 *PTPA* probe, two regions protected from DNase I digestion appeared (Figure 5, lanes 3–6). Both windows disappeared upon addition of an excess of unlabelled YY1-CON competitor (lanes 8 and 9), confirming the specificity of binding. Comparison with the G-specific sequencing reaction revealed that both protected regions contain the theoretical YY1 binding sites at nt −52/−44 and nt +27/+35, as expected. Thus these data confirm the presence of two YY1 binding sites in the *PTPA* promoter.

Functional analysis of the YY1 binding sites

To investigate the functional significance of the YY1 binding motifs, the −102/+38 YY1(u)mut-pLuc, −102/+38 YY1-(d)mut-pLuc and −102/+38 YY1(u,d)mut-pLuc plasmids were generated. In −102/+38 YY1(u)mut-pLuc and −102/+38 YY1(d)mut-pLuc, either the upstream (u) or the downstream (d) YY1 binding site is mutated: they contain the same two point mutations as described for PTPA-YY1(u) and PTPA-YY1(d) respectively. In −102/+38 YY1(u,d)mut-pLuc both YY1 binding sites are mutated.

First, these mutant *PTPA* promoter plasmids were tested in transient transfections in Saos-2, L929 and HepG2 cells in parallel with the wild-type −102/+38-pLuc construct (Figure 6). The results clearly demonstrate that mutation of the downstream YY1 binding site had minor effects on overall promoter activity (50–60% of promoter activity remaining). Conversely, promoter activity fell dramatically after mutation of the upstream YY1 binding site (90% decrease). This observation is in line with the results shown in Figure 2, where a similar decrease in promoter activity was seen upon deletion of the YY1(u) site

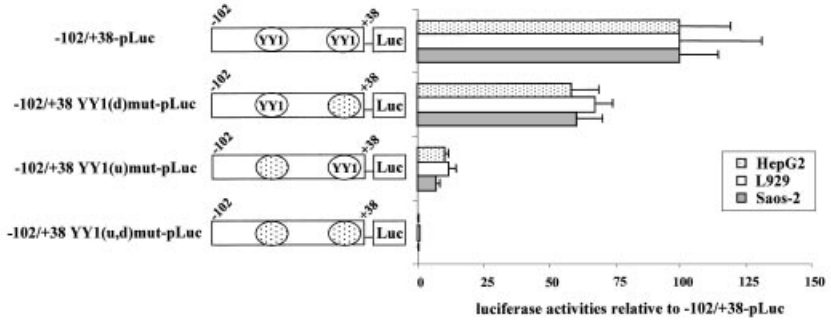

Figure 6 Functional analysis of both YY1 binding sites

Transient transfections were performed in HepG2, L929 and Saos-2 cells with the constructs indicated on the left: −102/+38-pLuc is the wild-type construct, whereas −102/+38 YY1(d)mut-pLuc, −102/+38 YY1(u)mut-pLuc and −102/+38 YY1(u,d)mut-pLuc contain two point mutations in the downstream (d), the upstream (u) or both YY1 binding motifs. Luciferase activities are presented as a percentage of the activity of −102/+38-pLuc, which was set at 100%. Means ± S.E.M. of at least three independent experiments are indicated.

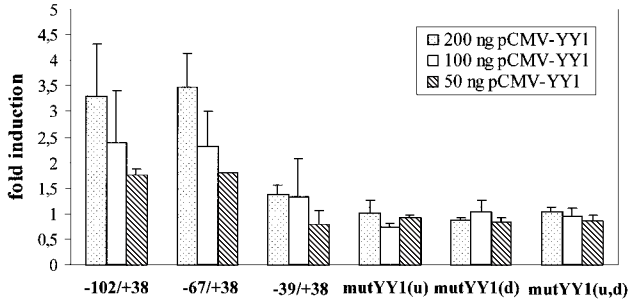

Figure 7 Co-transfection assays with pCMV-YY1 and *PTPA* promoter reporter plasmids in Saos-2 cells

Transient transfections were performed with 1 μg of −102/+38-pLuc, −67/+38-pLuc, −39/+38-pLuc, −102/+38 YY1(u)mut-pLuc, −102/+38 YY1(d)mut-pLuc or −102/+38 YY1(u,d)mut-pLuc, together with various amounts of pCMV-YY1 or empty pCMV (200, 100 and 50 ng). Induction factors were calculated as the ratios of luciferase activities of pCMV-YY1-co-transfected/pCMV-co-transfected reporter plasmids, minus the induction factor of the empty pGL2-Basic reporter plasmid (which maximally reached a value of 2). Means ± S.E.M. of three independent experiments are indicated.

(compare activities of −67/+38-pLuc and −39/+38-pLuc). Finally, it is clear that *PTPA* promoter activity is completely abolished if both YY1 binding sites are mutated.

In a second experiment, we performed co-transfections of −102/+38-pLuc, −67/+38-pLuc, −39/+38-pLuc and the three mutated versions of −102/+38-pLuc with a CMV-promoter-driven YY1 expression plasmid (pCMV-YY1, a gift from T. Shenk) in Saos-2 cells (Figure 7). Overexpression of YY1 resulted in induction of the observed promoter activities of −102/+38-pLuc and −67/+38-pLuc, whereas the activities of −39/+38-pLuc, −102/+38 YY1(u)mut-pLuc, −102/+38 YY1(d)mut-pLuc and −102/+38 YY1(u,d)mut-pLuc remained unaltered. Thus the YY1-inducibility of promoter activity clearly depends upon the presence of two intact YY1 binding sites (as in −102/+38-pLuc and −67/+38-pLuc), since deletion (as in −39/+38-pLuc) or mutation [as in −102/+38 YY1(u)mut-pLuc] of the upstream site, mutation of the downstream site [as in −102/+38 YY1(d)mut-pLuc] or mutation of both sites [as in −102/+38 YY1(u,d)mut-pLuc] results in loss of transcriptional stimulation upon pCMV-YY1 co-transfection. Moreover, the apparent stimulation of *PTPA* promoter activity by YY1 seems proportionate to the amount of co-transfected pCMV-YY1, demonstrating the dose-dependency of the effect (Figure 7). These results also indicate that, in Saos-2 cells, endogenous YY1 is not sufficient to saturate the introduced reporter constructs.

DISCUSSION

We have demonstrated that the human PTPA gene 5′ flanking sequence displays many characteristics of a TATA-less housekeeping gene promoter, including the presence of a 250 bp long unmethylated CpG island spanning part of exon I (nt −158/+91). In addition, the overall GC content of the promoter region is high (62% between nt −2681 and +210), reflected by the presence of many putative Sp1 binding sites.

In contrast with their established role in the regulation of many other housekeeping genes [25], these Sp1 binding sites seemed not to be required for basal expression of the PTPA gene. Indeed, the minimal *PTPA* promoter was localized between nt −67 and +38, a region that lacks Sp1 consensus sequences. Moreover, co-transfection of an Sp1 expression vector with −67/+38-pLuc in SL2 *Drosophila* Schneider cells, which lack endogenous Sp1 [30], did not result in any induction of the observed basal promoter activity (V. Janssens, unpublished work).

Instead, we demonstrated the functionality of two YY1 binding sites within the *PTPA* core promoter: one upstream site at positions −52/−44 and one downstream site at positions +27/+35. YY1 is a ubiquitously expressed multifunctional transcriptional regulator that can act as a repressor, an activator or an initiator element binding protein, depending on both the promoter and the cellular context (for a review, see [31]). Both repression and activation domains have been identified within YY1, which provides a structural basis for its dual functionality [32,33]. A large number of cellular and viral genes have been shown to be negatively regulated by YY1 (for reviews, see [31,34]). Examples of genes positively regulated by YY1 include c-*myc* [35], some ribosomal protein genes [36] and *p53* [37].

With regard to the PTPA gene, our data strongly support an activating role for YY1. Indeed, in transient transfection experiments in Saos-2 cells, it was demonstrated that overexpression of YY1 could enhance *PTPA* promoter activity by 3–4 times, depending on the amount of co-transfected pCMV-YY1. Moreover, this activation apparently relied on the integrity of both YY1 sites, since mutation of either one of them abolished transcriptional stimulation by YY1. In addition, it was shown that mutation of both YY1 binding sites, which destroyed YY1 binding *in vitro*, completely abolished promoter activity. Nevertheless, it seems that the distinct sites contribute differently to the observed basal promoter activity, since mutation (or deletion) of the upstream site alone resulted in a dramatic decrease in promoter activity, whereas mutation of the downstream site alone affected promoter activity only moderately. These differences in functionality might be explained in part by different binding affinities of YY1 for these sites *in vivo*.

YY1 may activate *PTPA* transcription by two mechanisms: (i) as an initiator protein, binding to the transcriptional start point region, or (ii) as a 'classical' transcription factor that modulates the basal transcription machinery by binding upstream and/or downstream of this initiator region. Considering that the unique *PTPA* transcription initiation site [15] is localized in between the two YY1 binding sequences, it seems more likely that YY1 activates the *PTPA* promoter by stimulating basal transcription and not by acting as an initiator.

Although YY1 is constitutively expressed, several mechanisms of direct or indirect regulation of YY1 activity have been published. Conflicting reports exist concerning the influence of CpG methylation on the binding capacity of YY1 to its cognate recognition site. Gaston and Fried [38] reported that binding of YY1 to the Surf-1/Surf-2 promoter is not affected by CpG methylation, whereas Satyamoorthy et al. [39] reported very poor binding of YY1 to CpG-methylated DNA. Since in the *PTPA* promoter the YY1 binding sites obviously reside within an unmethylated CpG island, this promoter may represent another interesting model to further clarify this issue.

In addition to these binding site modifications, the activity of YY1 can also be modulated by its interaction with other cellular proteins. The adenovirus E1A oncoprotein can both relieve YY1-mediated gene repression and enhance YY1-mediated gene activation, probably through indirect binding to YY1 via interaction with p300 [40]. The c-Myc protein can bind to YY1 as well, thereby inhibiting both its activating and its repressing abilities [41,42]. YY1 has also been shown to interact with Sp1 [43,44]. In the future it may be interesting to investigate whether these YY1-interacting and -regulating proteins have any influence on PTPA gene expression. Since many Sp1 sites are present upstream of the minimal *PTPA* promoter, it may be particularly

interesting to further explore the putative interplay between Sp1 and YY1 in the context of this housekeeping gene promoter.

We thank Fabienne Withof, Maria Veeckmans and Robin Justé for expert technical assistance, and Valère Feytons for oligonucleotide synthesis. Thomas Shenk (Howard Hughes Medical Institute and Department of Molecular Biology, Princeton University, Princeton, NJ, U.S.A.) is acknowledged for the gift of the YY1 expression plasmid. This work was supported by the Fonds voor Wetenschappelijk Onderzoek Vlaanderen, the Geconcerteerde OnderzoekActies van de Vlaamse Gemeenschap, the Human Frontier Science Program and the E.C. Cancer Program Biomed2. C.V.H. is the holder of a postdoctoral fellowship from the FWO Vlaanderen.

REFERENCES

1 Mumby, M. C. and Walter, G. (1993) Physiol. Rev. **73**, 673–699
2 Mayer-Jaekel, R. E. and Hemmings, B. A. (1994) Trends Cell Biol. **4**, 287–291
3 Chernoff, J., Li, H.-C., Cheng, Y.-S. E. and Chen, L. B. (1983) J. Biol. Chem. **258**, 7852–7857
4 Silberman, S. R., Speth, M., Nemani, R., Ganapathi, M. K., Dombradi, K., Paris, H. and Lee, E. Y. C. (1984) J. Biol. Chem. **259**, 2913–2922
5 Cayla, X., Ballmer-Hofer, K., Merlevede, W. and Goris, J. (1993) Eur. J. Biochem. **214**, 281–286
6 Chen, J., Martin, B. L. and Brautigan, D. L. (1992) Science **257**, 1261–1264
7 Brautigan, D. L., Chen, J. and Thompson, P. (1993) Adv. Protein Phosphatases **7**, 49–65
8 Chen, J., Parsons, S. and Brautigan, D. L. (1994) J. Biol. Chem. **269**, 7957–7962
9 Cayla, X., Goris, J., Hermann, J., Hendrix, P., Ozon, R. and Merlevede, W. (1990) Biochemistry **29**, 658–667
10 Van Hoof, C., Cayla, X., Bosch, M., Merlevede, W. and Goris, J. (1994) Adv. Protein Phosphatases **8**, 301–330
11 Janssens, V., Van Hoof, C., Goris, J. and Merlevede, W. (1998) Methods Mol. Biol. **93**, 103–115
12 Cayla, X., Van Hoof, C., Bosch, M., Waelkens, E., Vandekerckhove, J., Peeters, B., Merlevede, W. and Goris, J. (1994) J. Biol. Chem. **269**, 15668–15675
13 Van Hoof, C., Janssens, V., Dinishiotu, A., Merlevede, W. and Goris, J. (1998) Biochemistry **37**, 12899–12908
14 Van Hoof, C., Cayla, X., Bosch, M., Merlevede, W. and Goris, J. (1994) Eur. J. Biochem. **226**, 899–907
15 Van Hoof, C., Sayed Aly, M., Garcia, A., Cayla, X., Cassiman, J. J., Merlevede, W. and Goris, J. (1995) Genomics **28**, 261–272
16 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
17 Majumdar, D., Avissar, Y. J. and Wyche, J. H. (1991) Biotechniques **11**, 94–101
18 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
19 Dignam, J. D., Lebovitz, R. L. and Roeder, R. G. (1983) Nucleic Acids Res. **11**, 1475–1489
20 Park, K. and Atchison, M. L. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 9804–9808
21 Courtois, S. J., Lafontaine, D. L., Lemaigre, F. P., Durviaux, S. M. and Rousseau, G. G. (1990) Nucleic Acids Res. **18**, 57–64
22 Dutton, C. M., Paynton, C. and Sommer, S. S. (1993) Nucleic Acids Res. **21**, 2953–2954
23 Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. **65**, 499–560
24 Smale, S. T. (1997) Biochim. Biophys. Acta **1351**, 73–88
25 Azizkhan, J. C., Jensens, D. E., Pierce, A. J. and Wade, M. (1993) Crit. Rev. Eukaryotic Gene Expression **3**, 229–254
26 Gardiner-Garden, M. and Frommer, M. (1987) J. Mol. Biol. **196**, 261–282
27 Bird, A. P. (1992) Cell **70**, 5–8
28 Herman, J. G., Latif, F., Weng, Y., Lerman, M. I., Zbar, B., Liu, S., Samid, D., Duan, D. S., Gnarra, J. R. and Linchan, W. M. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 9700–9704
29 Hyde-DeRuyscher, R. P., Jennings, E. and Shenk, T. (1995) Nucleic Acids Res. **23**, 4457–4465
30 Courey, A. J. and Tjian, R. (1988) Cell **55**, 887–898
31 Shi, Y., Lee, J.-S. and Galvin, K. M. (1997) Biochim. Biophys. Acta **1332**, F49–F66
32 Lee, T.-C., Zhang, Y. and Schwartz, R. J. (1994) Oncogene **9**, 1047–1052
33 Lee, J.-S., See, R. H., Galvin, K. M., Wang, J. and Shi, Y. (1995) Nucleic Acids Res. **23**, 925–931
34 Shrivastava, A. and Calame, K. (1994) Nucleic Acids Res. **22**, 5151–5155
35 Riggs, K. J., Saleque, S., Wong, K.-K., Merrell, K. T., Lee, J.-S., Shi, Y. and Calame, K. (1993) Mol. Cell. Biol. **13**, 7487–7495
36 Hariharan, N., Kelley, D. and Perry, R. P. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 9799–9803
37 Furlong, E. E. M., Rein, T. and Martin, F. (1996) Mol. Cell. Biol. **16**, 5933–5945
38 Gaston, K. and Fried, M. (1995) Nucleic Acids Res. **23**, 901–909
39 Satyamoorthy, K., Park, K., Atchison, M. L. and Howe, C. C. (1993) Mol. Cell. Biol. **13**, 6621–6628
40 Lee, J.-S., Galvin, K. M., See, R. H., Eckner, R., Livingston, D., Moran, E. and Shi, Y. (1995) Genes Dev. **9**, 1188–1198
41 Shrivastava, A., Saleque, S., Kalpana, G. V., Artandi, S., Goff, S. P. and Calame, K. (1993) Science **262**, 1889–1892
42 Shrivastava, A., Yu, J., Artandi, S. and Calame, K. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 10638–10641
43 Seto, E., Lewis, B. and Shenk, T. (1993) Nature (London) **365**, 462–464
44 Lee, J.-S., Galvin, K. M. and Shi, Y. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 6145–6149

Received 28 May 1999/19 August 1999; accepted 1 October 1999