Functional characterization of transcriptional regulatory elements in the upstream region and intron 1 of the human S6 ribosomal protein gene

Marianne ANTOINE and Paul KIEFER[1]

Abteilung für Virologie, Ruhr-Universität Bochum, Geb. MA 6/130, D-44780 Bochum, Germany

Expression of housekeeping genes involves regulation at comparable levels in a wide spectrum of cells. To define the *cis*-regulatory elements in the human S6 ribosomal protein (rpS6) gene, we made a series of deletions of the upstream non-transcribed region, including or excluding exon 1 or intron 1 sequences. The mutated rpS6 gene regulatory regions were fused to the chloramphenicol acetyltransferase reporter gene and transfected into HeLa and COS-1 cells. The results have identified three parts of the rpS6 gene that are required for efficient and specific transcription. The core promoter includes only a 40 bp region upstream of the transcription start site and initiation region. Both upstream and intronic elements enhance transcription from the core promoter. Furthermore, mutation of the splice donor site of intron 1 almost completely abolished the enhancing activity of the intronic transcriptional modulator. We used gel retardation assays to identify sequence-specific binding sites in the upstream region and in the proximal half of intron 1. Both common and different nuclear factors that bind the rpS6 gene promoter were identified in extracts from HeLa and COS-1 cells, suggesting that different transcription factors may bind specifically to the same binding region and might be interchangeable in their function to ensure high-level expression of housekeeping genes independently of the cell type.

INTRODUCTION

The formation of ribosomes, which is a prerequisite for protein synthesis (e.g. during G1 progression), requires the co-ordinated expression of over 80 different ribosomal proteins. In proliferating cells, the steady-state levels of different ribosomal protein mRNAs are remarkably uniform, although during differentiation the transcription of ribosomal protein genes is co-ordinately down-regulated [1–5]. Since ribosomal protein genes are not found clustered in the mammalian genome, their uniform expression is rather the result of common *trans*-acting transcription factors and/or post-transcriptional regulation.

Comparisons of the mammalian ribosomal protein transcriptional promoters analysed to date reveal a number of common features. These include the absence of a TATA box motif, location in a CpG island and the location of the transcription initiation point(s) at C residues embedded within a polypyrimidine tract. In several mouse ribosomal protein genes, binding motifs for ubiquitously expressed transcription factors have been identified located upstream and/or downstream of the transcription initiation point(s). The *cis*-acting binding elements present in the regulatory regions of the L30 ribosomal protein (rpL30), rpL32 and rpS16 genes are quite similar, yet the *trans*-acting factors that actually bind to particular regulatory elements of rpS16 appear to be different from those interacting with the corresponding elements of rpL30 and rpL32 [6–14].

We are interested in the organization and expression of the S6 ribosomal protein (rpS6) gene, which encodes the major target for phosphorylation by serine kinases in the 40 S ribosomal subunit. Phosphorylation of rpS6 has been demonstrated on stimulation of a variety of cells by mitogenic agents and oncogene products, and is associated with facilitating the translation initiation and selective translation of certain mRNAs [15]. In an earlier report, we described the organization and chromosomal localization of the human rpS6 gene. This gene exhibits typical features of a ribosomal protein gene, including a small first exon, a short untranslated leader sequence, the transcription start sites at cytidine residues embedded in a polypyrimidine tract, the absence of a TATA box and location of the 5′ end within a CpG-rich island [16].

In the present study we have analysed the transcriptional promoter region of the human rpS6 gene. The effects of various 5′ deletion mutants and of the deletion of various regions of the first intron on the transcriptional regulation of the rpS6 gene have been evaluated in transfected HeLa cells and COS-1 cells. Our findings reveal that, although the contribution of various elements to rpS6 promoter activity is similar in both cell types, clear differences are detected in the binding of nuclear factors to corresponding elements. Some of the differences, detected in the gel mobility shift experiments, showed a clear correlation with differences in the promoter activity of deletion mutants. These results indicate that, in housekeeping genes such as ribosomal protein genes, regulatory elements may be abundant and some transcription factors binding to such sites may be interchangeable in order to achieve similar levels of expression in different cell types.

EXPERIMENTAL

Cell culture, DNA transfection, chloramphenicol acetyltransferase (CAT) assays and β-galactosidase (β-GAL) assays

COS-1 cells and HeLa cells were transfected with 20 μg of test plasmid and 5–10 μg of pJ7 lacZ (β-GAL) by electroporation (270 V/960 μF for HeLa cells and 450 V/250 μF for COS-1 cells) in PBS. At 48 h after transfection the cells were harvested and resuspended in 250 mM Tris/HCl, pH 8.0, and cell extracts were prepared for CAT and β-GAL assays by three cycles of freeze/thawing as recommended by the suppliers of the assay kits (Promega).

Abbreviations used: rpS6 (etc.), S6 ribosomal protein (etc.); CAT, chloramphenicol acetyltransferase; β-GAL, β-galactosidase; YY1, Yin-Yang 1; GABP, GA binding protein.

[1] To whom correspondence should be addressed (e-mail Paul.Kiefer@rz.ruhr-uni-bochum.de).

β-GAL activity was used as an internal reference of transfection efficiency, and was assayed photometrically. The standard assay was performed by adding a diluted extract sample to an assay buffer (final concentrations: 0.1 M sodium phosphate buffer, pH 7.5, 1 mM $MgCl_2$, 45 mM β-mercaptoethanol) containing *o*-nitrophenyl β-D-galactopyranoside (0.8 mg/ml). The reactions were incubated at 37 °C for 30 min, terminated, and the absorbance at 420 nm was read. The readings were corrected using blank values obtained from extracts from mock transfections.

The extracts were assayed for CAT enzyme activity after incubation at 65 °C for 10 min to inactivate endogenous acetylase. The cell extracts were tested for acetylation of ^{14}C-labelled chloramphenicol (at 0.025 mCi/ml; Amersham) in the presence of n-butyryl-CoA (at 5 mg/ml; Sigma). The n-butyryl chloramphenicol partitions were extracted into the xylene phase and, after two brief back-extractions, a portion of the xylene phase was mixed with scintillant and radioactivity was counted in a scintillation counter. A standard curve of CAT activity was prepared using a blank from mock-transfected cell extracts and CAT enzyme controls to ensure linearity of the enzyme reaction. The CAT values of each extract tested were corrected by taking account of the β-GAL activity.

RNase protection assay

Total RNA was isolated from transiently transfected HeLa or COS-1 cells on a CsCl/trifluoroacetate gradient as recommended by Pharmacia. Antisense ^{32}P-labelled RNA probes were synthesized from DNA plasmids using SP6 polymerase under conditions recommended by the supplier (Promega). Probe 1 was prepared from plasmid CATRI S61, which contained nucleotides -46 to $+48$ (46 nucleotides of the 5′ region and exon 1) and nucleotides 569–594 (part of exon 2) of rpS6, plus 268 nucleotides of the CAT gene coding region, cloned in the vector pGEM4Z [16]. Probe 2 was prepared from plasmid CATRI S62, which contained nucleotides -5 to $+594$ (including intron 1) of rpS6 plus 268 nucleotides of the CAT gene coding region cloned in the vector pGEM4Z. Probe 3 was prepared from plasmid CATRI S63, containing nucleotides -46 to $+48$, 701 to 1269 (intron 2) and 1269 to 1277 (part of exon 3) of rpS6, plus 268 nucleotides of the CAT gene coding region. For RNA synthesis, the plasmids were linearized at a *Hin*dIII site (multicloning site of pGEM4Z). The RNA was analysed by RNase protection as described [17].

Preparation of nuclear extracts and gel retardation assays

Nuclear extracts were prepared from HeLa cells and COS-1 cells as described by Wildeman et al. [18]. Protein concentrations were determined using the Bio-Rad protein assay procedure. Single-stranded oligonucleotide probes were annealed to the respective complementary oligonucleotide and labelled (100 ng) with polynucleotide kinase. Unincorporated label was removed by Sephadex G25 column chromatography. Between 10000 and 25000 c.p.m. of each probe (0.25 ng) was incubated with 10 μg of HeLa or COS-1 cell extract and 3 μg of poly(dA-dT) or poly(dG-dC) in binding buffer (20 mM Tris, pH 8, 5 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM EDTA, 60 mM KCl) at room temperature for 20 min. Complexes were resolved on non-denaturing 5% (w/v) polyacrylamide gels run in 0.5 × Tris/borate/EDTA buffer at room temperature.

Plasmid constructions

rpS6–CAT fusion plasmids

These were constructed by replacing sequences between the *Hin*dIII and *Xba*I sites of the pCAT basic vector (Promega) with rpS6 sequences generated by PCR. pCAT6.1 was constructed using PCR with a sense primer around a *Spe*I restriction site (-876) including an additional *Hin*dIII site and an antisense primer including the first 24 bases of exon 2 plus a *Xba*I site. We used as a template the previously described genomic clone pS6EcoRI, which covers about 3.5 kb of the 5′ non-transcribed region and exons 1 and 2, including intron 1 [16]. pCAT6.3 was constructed by PCR with the same primer set but using as template a plasmid in which an rpS6 genomic *Spe*I (-876) to *Pvu*II ($+32$) DNA fragment was joined through the *Pvu*II site to the corresponding rpS6 cDNA fragment. Plasmids pCAT6.2 and pCAT6.4 were derived from plasmids pCAT6.1 and pCAT6.3 repectively by replacing a *Pvu*II/*Xba*I fragment with the corresponding double-stranded oligonucleotide, introducing a point mutation to change the translation start codon ATG into ATC. Plasmid pCAT6.5 was constructed by inserting the *Hin*dIII/*Pvu*II fragment of pCAT6.3 between the *Hin*dIII and *Acc*I sites in the pCAT basic vector.

5′ deletion mutants

A set of progressive 5′ deletion mutants was created by PCR with selected 5′ primers (as indicated in Figure 2) which included a *Hin*dIII site, and a common 3′ primer which included the unique *Bam*HI site ($+136$) present in the first rpS6 intron. These PCR products were ligated into *Hin*dIII/*Bam*HI-digested pCAT6.2 and successively numbered pCAT6.21–pCAT6.26.

Intron and splice site mutants

The 3′ deletion construct pCAT6.6 was generated by cloning a 635 bp PCR product into *Bam*HI/*Nco*I-digested pCAT6.2, deleting the intronic sequences between nt 141 and 527 of the rpS6 gene. The PCR product included 42 nt from the 3′ end of intron 1 and a new *Bam*HI cloning site; the 3′ primer was derived from the region around the unique *Nco*I site present in the CAT coding sequence. A second 3′ deletion construct, pCAT6.7, was generated by fusing a 234 bp PCR DNA fragment amplified from the pCAT6.2 region spanning nt -46 (including the single *Nru*I site) to nt $+136$ (*Bam*HI site) to the 635 bp *Bam*HI/*Nco*I PCR product described above, and cloning into *Nru*I/*Nco*I-digested pCAT6.2. In order to destroy the donor splice site of intron 1, a PCR product was generated to mutate the consensus sequence AG:gtaggt into AG:ctcgag, and the mutated DNA fragment was inserted into pCAT6.2 to generate pCAT6.8. The plasmid pCAT6.9 was constructed by ligating a 908 bp *Spe*I/*Pvu*II fragment of pCAT6.3 to intron 2 sequences generated by PCR and cloning the product into *Spe*I/*Xba*I-digested pCAT6.2. The 5′ primer contained 13 bases of exon 1 (nucleotides 36–48), including the *Pvu*II cloning site and the proximal 25 nt of intron 2. The 3′ primer contained a 3′ *Xba*I cloning site, the proximal 9 bases of exon 3 (nt 1269–1277) and 14 nucleotides of intron 2 (nt 1268–1255) for specific priming. The exon/intron boundaries should ensure correct processing of the intron 2 sequences. Where possible, DNA fragments derived from PCR products were replaced by the corresponding genomic sequences, and all plasmid constructs were verified by DNA sequencing.

RESULTS

Sequences important for efficient promoter activity

In order to identify functionally important regions of the human rpS6 gene, a number of rpS6–CAT chimaeras were prepared by linking various parts of the rpS6 gene to a promoterless CAT gene. After transfection into COS-1 or HeLa cells, the transcriptional activity of the constructs was assessed by measuring

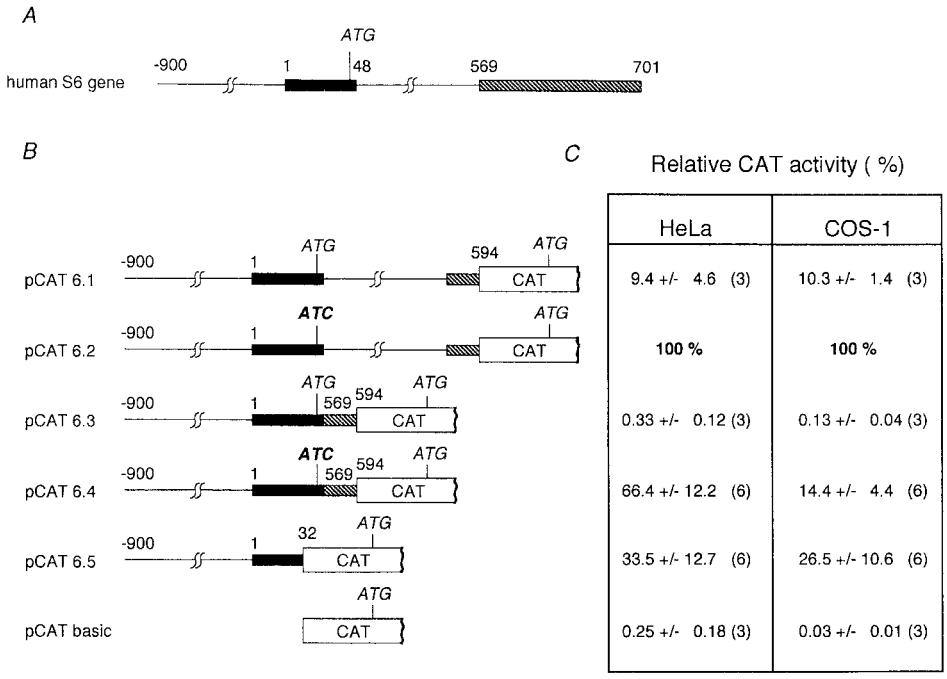


Figure 1 rpS6 promoter and influence of intron 1 sequences on transcription

(**A**) Structure of the 5′ region of the rpS6 gene showing exons 1 and 2 (black and hatched boxes respectively), intron 1 and 900 bp of the 5′ upstream region. (**B**) Schematically depicted structure of rpS6–CAT fusion plasmids. The position of the translation initiation site in exon 1 is marked. The intronic sequences and the 3′ upstream region are represented as thin lines, while exon 1 is depicted as a black box and exon 2 as a hatched box. (**C**) The values of the CAT activities were normalized to β-GAL activity as an internal reference of the transfection efficiency, and are expressed as a percentage of the CAT activity of the pCAT6.2 plasmid (taken as 100%). The data are expressed as means $\pm$ S.D., with the numbers of independent transfections in parentheses.

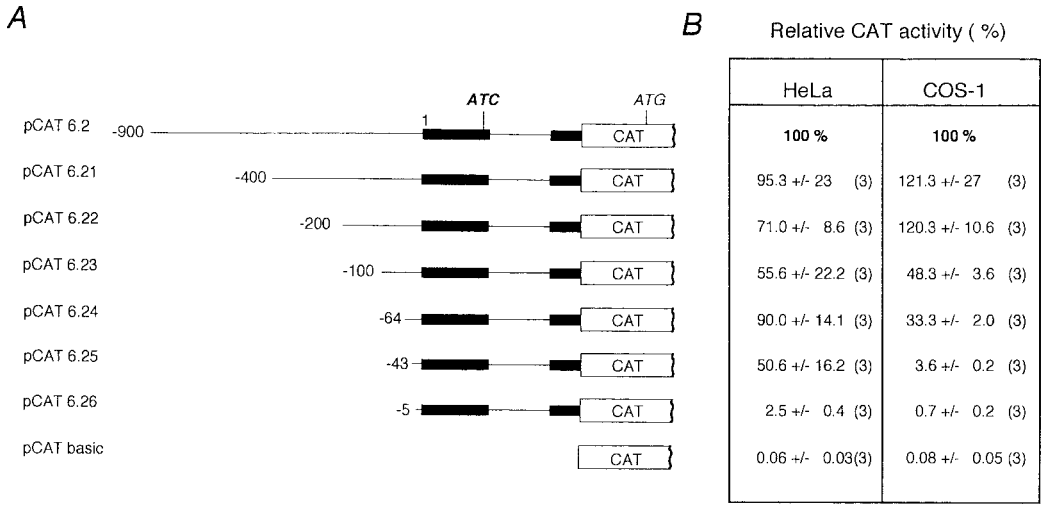


Figure 2 Deletion analysis of the rpS6 5′ upstream region in the presence of intron 1

(**A**) Schematic diagram depicting the extent of 5′ deletion in a set of mutants based on plasmid pCAT6.2. (**B**) Relative levels of CAT expression in cell extracts from HeLa and COS-1 cells transfected with the various deletion mutants (see the legend to Figure 1 for details).

CAT activity in the cell extracts in comparison with a co-transfected β-GAL control. Initially, we constructed rpS6–CAT chimaeras containing about 1 kb of the non-transcribed 5′ region and combinations of exon 1, intron 1 and 5′ sequences of exon 2 to ensure correct splicing, as illustrated schematically in Figure 1. On average, construct pCAT6.2 showed the highest level of promoter activity, which was designated 100% relative CAT activity (Figure 1). The ATG codon in exon 1, which normally initiates rpS6 protein synthesis, severely represses CAT activity, since its conversion into ATC increased CAT levels 10-fold (Figure 1; compare pCAT6.1 with pCAT6.2). An enhancement of CAT activity was also apparent if the intervening sequences encompassing the ATG codon and intron 1 were deleted (Figure 1; pCAT6.5). A comparison of similar constructs lacking intron

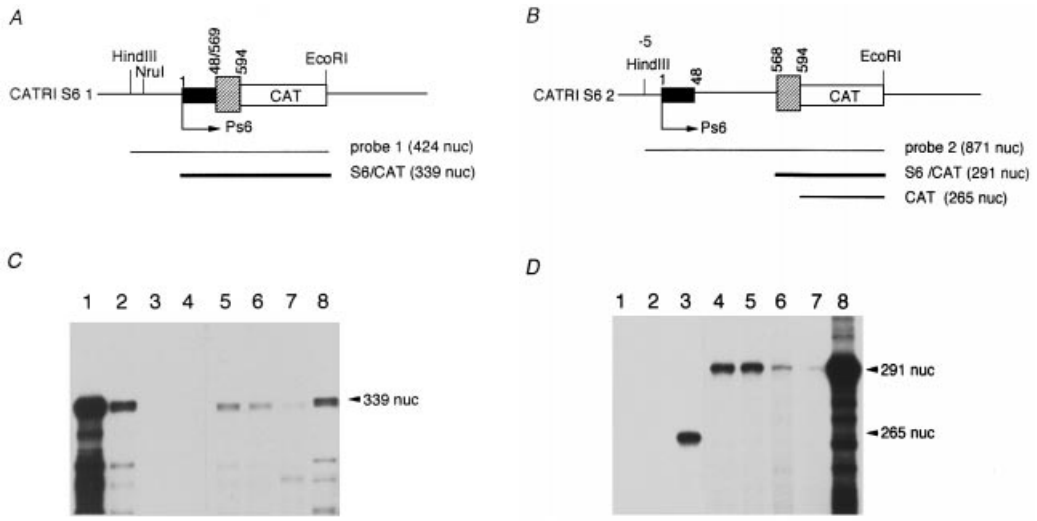


Figure 3 Promoter efficiency, cap site usage and processing efficiency of the pCAT6.2 deletion mutants determined by RNase protection assay

Schematic diagrams of CATRI S61 (**A**) and CATRI S62 (**B**), derived from pCAT6.4 and pCAT6.2 respectively, are shown in the upper panels. The locations of the antisense RNA probes (1 and 2) are depicted as thin lines, and the predicted sizes of the protected fragments [in nt ('nuc')] are indicated as thickened lines. Autoradiographs showing the protected fragments obtained with probe 1 (**C**) and probe 2 (**D**) are shown in the lower panels. Total RNA from HeLa cells transiently transfected with the indicated deletion mutants was analysed. The sizes of the protected fragments were calculated by comparison with a known sequencing ladder run in parallel. Constructs: pCAT6.2 (lanes **C**1 and **D**8), pCAT6.4 (lane **C**2), CAT basic (lanes **C**4 and **D**2), pCAT6.21 (lanes **C**5 and **D**4), pCAT6.24 (lanes **C**6 and **D**5), pCAT6.25 (lanes **C**7 and **D**6), pCAT6.26 (lanes **C**8 and **D**7). Controls were extracts from non-transfected HeLa cells (lanes **C**3 and **D**1) and cells transfected with pCAT basic (lanes **C**4 and **D**2) as negative controls, and from cells transfected with pCAT control (lane **D**3) as a positive control.

1 with and without the ATG codon confirmed the importance of removing the ATG start site for rpS6 synthesis. This comparison also indicated a cell-type difference in transcriptional activity between fibroblast-like COS-1 and epithelial-like HeLa cells, since COS-1 cells required intron 1 in order to show activity comparable with that of HeLa cells (Figure 1; compare pCAT6.4 with pCAT6.2). Although transcription factors should be highly conserved between human cells and monkey cells, these differences may also be species-related. However, intron 1 sequences placed in different orientations with respect to a simian virus 40 promoter did not show similar differential effects, suggesting that they do not function as a classical enhancer (results not shown).

The rpS6 core promoter is contained within a region of fewer than 43 bp upstream of the transcription start site

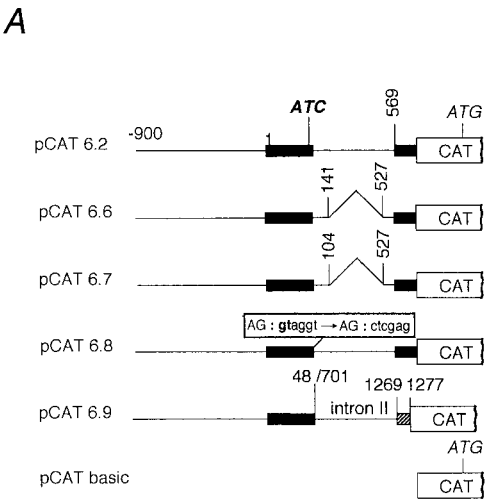
To identify the 5′ flanking regions that regulate rpS6 gene expression, these sequences in plasmid pCAT6.2 were subjected to a progressive deletion analysis (Figure 2). A deletion from nt −900 to nt −200 (from the 5′ end of pCAT6.2) had only a modest effect on CAT activity; removal of a further 100 nt reduced CAT activity in both cell lines by approx. 50%. However, deletion to nt −43 had little effect on the CAT activity in HeLa cells, but caused more than 90% loss of the original activity in COS-1 cells, suggesting that different transcription factors are utilized in these two cell lines. Deletion to nt −5 essentially abolished CAT activity in both cell lines. To ensure comparability between experiments, a complete set of mutant plasmids was transfected and the CAT activity standardized to the internal standard provided by β-GAL activity, as described.

To assess whether the presence of intron 1 sequences may influence the requirement for 5′ upstream sequences for transcriptional activity, plasmid pCAT6.5 (which is similar to pCAT6.2, but lacks the intron 1 region) was subject to a similar 5′ sequence deletion analysis and the resultant plasmids were tested in COS-1 and HeLa cells. For each mutant, a similar decrease in CAT activity was observed for both cell lines; in particular, the construct containing only 5 nt upstream gave no detectable CAT activity. Interestingly, the deletion mutant with a 5′ region deletion corresponding to mutant pCAT6.24 (see Figure 2) did not show the higher activity associated with HeLa cells, but rather the lower activity reminiscent of COS-1 cells. Likewise, deletion of the region between nt −64 and −43 resulted in a sharp decrease in CAT activity to less than 5% of that of the parent plasmid in both cell lines (results not shown).

To demonstrate that the observed CAT activities are based on correctly initiated and spliced transcripts, total RNA isolated from transfected HeLa and COS-1 cells was analysed by RNase protection assay (Figure 3). The amounts of transcript derived for each mutant were measured using ^{32}P-labelled antisense RNA probes which encompassed the polypyrimidine tract and 46 bp of upstream rpS6 sequence. Probe 1 was derived from pCATRI S61, and probe 2 (which includes intron 1) was from pCATRI S62. The results show that protected fragments corresponded to transcripts initiated correctly at the polypyrimidine tract, as described for the rpS6 gene, and that intron 1 had been removed. We conclude from these results that, in the absence of intron 1 sequences, 40 bp of the 5′ non-transcribed region are sufficient for correct initiation at the polypyrimidine tract; however, in the presence of intron 1, as few as 5 bp upstream results in correctly initiated transcripts in HeLa cells.

Major requirements for transcription enhancement by intron 1 of rpS6

To investigate the effect of intron 1 on the transcription of the rpS6 gene, we generated a new series of plasmids in which parts of intron 1 were deleted or exchanged for intron 2, including 5′ sequences of exon 3 to facilitate splicing (Figure 4A). In another



B

Relative CAT activity (%)

HeLa	COS-1
100 %	100 %
22.3 +/- 5.7 (3)	19.5 +/- 1.7 (3)
37.3 +/- 7.4 (3)	25.0 +/- 2.4 (3)
0.5 +/- 0.1 (3)	3.3 +/- 0.3 (3)
6.4 +/- 2.2 (3)	4.1 +/- 0.23 (3)
0.21 +/- 0.15 (3)	0.01 +/- 0.004 (3)

Figure 4 Effect of the presence of intron 1 of rpS6 on CAT activity

(**A**) A series of mutant cDNAs, all derivatives of pCAT6.2, are depicted schematically. In the constructs pCAT6.6 and pCAT6.7 the intron sequences are deleted between nt 141 and 527 and between nt 104 and 527 respectively. In the mutant pCAT6.8 the donor splice site is mutated, and in pCAT6.9 intron 1 is replaced by intron 2. (**B**) The different plasmids were introduced into HeLa and COS-1 cells by electroporation, and 48 h later the relative CAT activities were measured (see the legend to Figure 1 for details).

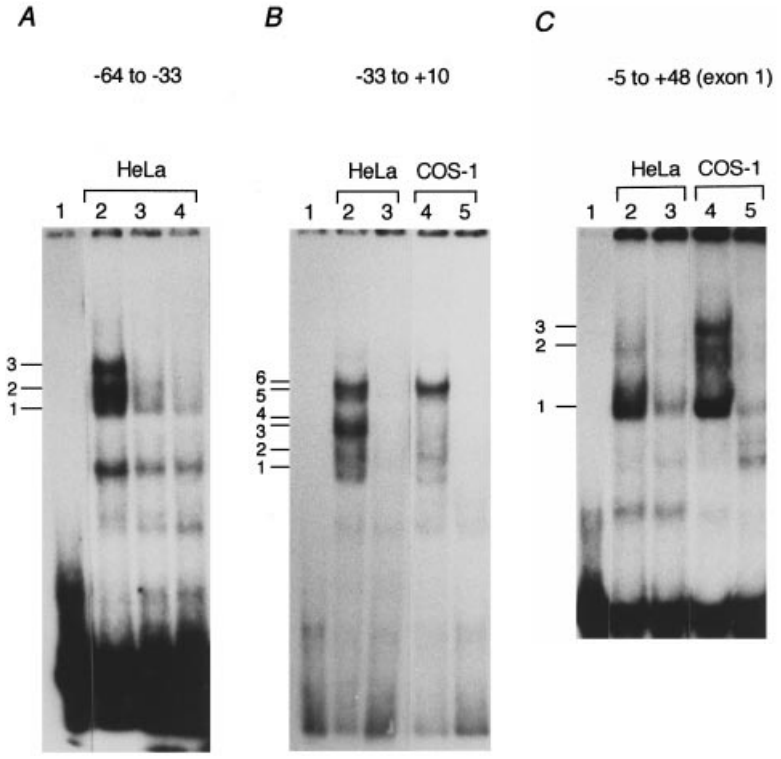


Figure 5 Common and cell-type-specific nuclear factors binding to the 5′ upstream region and exon 1 of the rpS6 gene determined by gel mobility shift assay

(**A**) ^{32}P-labelled fragments (nt −64 to −33) were incubated in the absence (lane 1) or the presence (lanes 2–4) of nuclear extracts isolated from HeLa cells. Lanes 3 and 4 show the effects of competition with unlabelled oligonucleotide −64 to −33 (50 ng and 100 ng respectively). Poly(dG-dC) was used as a non-specific competitor. (**B**) ^{32}P-labelled fragment −33 to +10 was incubated in the absence (lane 1) or the presence of HeLa cell extracts (lanes 2 and 3) or COS-1 cell extracts (lanes 4 and 5). The effects of specific competition (unlabelled oligonucleotide −33 to +10) are shown in lanes 3 and 5. In these assays, poly(dA-dT) was used as a non-specific competitor. (**C**) ^{32}P-labelled oligonucleotide −5 to +48, encompassing exon 1, was incubated in the absence (lane 1) or the presence of HeLa cell extracts (lanes 2 and 3) or COS-1 cell extracts (lanes 4 and 5). Lanes 3 and 5 show the effect of competition with 50 ng of unlabelled oligonucleotide. Poly(dG-dC) was used as a non-specific competitor.

construct, the splice donor site of intron 1 was mutated to determine whether the intron had to be removed in order to enhance rpS6 transcription. The mutant constructs were transiently transfected into COS-1 and HeLa cells, and the resultant CAT activities were shown to be comparable for the two cell lines. Deletion of the intronic sequences between nt 104 and 527 or between nt 141 and 527 resulted in a decrease in CAT activity to less than 25% of the values obtained with the parent plasmid pCAT6.2 (Figure 4B). An almost complete loss of activity was recorded when the donor splice site was mutated. As expected from results with the deletion mutants, replacement of intron 1 with intron 2 sequences could not restore CAT activity. However, this mutant showed a significantly higher activity than the construct with the non-functional splice site. RNase protection assays showed that construct pCAT6.9 gave correctly processed transcripts (results not shown). Taken together, the data indicate that the effect of intron 1 is sequence-specific, and in order to gain full transcriptional activity the intron has to be removed.

Gel retardation analysis of the core promoter and exon 1 region of the human rpS6 gene

Given the array of functional elements and the obvious differences between COS-1 and HeLa cells, it was of considerable interest to determine whether the locations of deletions causing a change in transcriptional activity of the rpS6 gene correlated with the position of candidate binding sites for nuclear factors. Gel retardation analysis was used to identify the binding of nuclear factors to rpS6 sequences present in nuclear extracts prepared from both COS-1 cells and HeLa cells. Since we were particularly interested in identifying factors which may be cell (type)-specific or species-related, we focused our search for binding sites on the region encompassing 64 nt of the 5′ non-transcribed region, the first exon and the first 52 nt of intron 1. Our gel retardation assay with a ^{32}P-labelled oligonucleotide extending between nt -64 and -33 of rpS6 demonstrated at least three nuclear factors binding to this region in both COS-1 and HeLa cells. However, we found that the choice of the non-specific competitor had a profound effect on the binding of nuclear factors to this oligonucleotide. Apart from three highly retarded bands (Figure 5A, bands 1–3 in lane 2) found with nuclear extracts from both cell lines, an additional band specific to COS-1 cell nuclear extracts was observed when using poly(dA-dT) as non-specific competitor. The specificity of these bands was demonstrated by competition experiments with an excess of unlabelled oligonucleotides (Figure 5A, lanes 3 and 4).

Because the -43 deletion mutant showed a clear difference in CAT activity when transfected into COS-1 cells or HeLa cells, we expected that an oligonucleotide extending from nucleotide -33 to $+10$ may have a cell-specific factor binding pattern. In fact, three band shifts are common to both cell lines (bands 1, 2 and 6, lanes 2 and 4, Figure 5B) and three additional retarded bands are observed with HeLa extracts (bands 4, 5 and 6, lane 2, Figure 5B). All these bands were shown to be specific as they were successfully competed by the unlabelled oligonucleotide (lanes 3 and 5, Figure 5B). No difference was noted with the band shift pattern of the -33 to $+10$ oligonucleotide when using either poly(dG-dC) or poly (dA-dT) as a non-specific competitor. When poly(dI-dC) was included in the assay as a non-specific competitor some factor binding was lost with extracts from both cell lines. Interestingly a specific competitor encompassing nt -39 to -15 overlapping with oligonucleotide -33 to $+10$ by 19 nt abolishes a major band using COS-1 cell extracts, and less convincingly a band using HeLa cell extracts, suggesting that at least one factor binds to the 5′ half of the oligonucleotide -33 to $+10$ (results not shown).

Figure 6 Oligonucleotide -19 to $+1$ encompasses a possible ETS binding site

HeLa cell extracts (lanes 2 and 4) or COS-1 cell extracts (lanes 3 and 5) were incubated with labelled oligonucleotide -19 to $+1$, which includes an ETS factor binding consensus sequence (GGAA). A 50 ng portion of unlabelled competitor was enough to abolish protein–DNA complex-formation (lanes 4 and 5). Lane 1 shows the labelled oligonucleotide incubated without protein extract. Poly(dG-dC) was used as a non-specific competitor.

Using an oligonucleotide from nt -5 to nt $+48$, which encompasses all of exon 1, a major fast-migrating band and one weaker slow-migrating band were detected in extracts from both cell types (Figure 5C, bands 1 and 2 in lanes 2 and 4). In addition, a slower-migrating band was detected only with COS-1 cell extracts (Figure 5C, band 3 in lane 4). No differences in the band-shift pattern were detected using an oligonucleotide in which the ATG codon was mutated to ATC.

When an oligonucleotide encompassing nt -19 to $+1$ was used in a gel mobility shift assay, a single retarded band was revealed using COS-1 extracts, and an equally intense second band was revealed using HeLa extracts (Figure 6). Interestingly, this oligonucleotide probe contains a core consensus sequence, GGAA, for the ETS family of transcription factors.

Recombinant YY1 (Yin-Yang 1) protein binds to the rpS6 intronic transcription modulator

Overlapping oligonucleotides derived from the 5′ region of the first intron were also tested for their abiltiy to bind nuclear factors from HeLa and COS-1 cells. In Figure 7 the results are presented from this analysis using HeLa cell extracts and oligonucleotides from nt 68 to nt 47, which includes a YY1 consensus binding site (GCCACCATCAC), from nt 40 to nt 82 and from nt 70 to nt 100. At least three sets of the multiple band shifts observed using the nt 40–82 oligonucleotides were common to the nuclear extracts from both cell types (Figure 7A). However, depending on the non-specific competitor used, some of the minor bands were detected exclusively with one or the other cell extract.

The nt 68–47 oligonucleotide revealed one shifted band migrating in a similar position to a major band detected with the nt 40–82 fragment, suggesting that the binding of additional

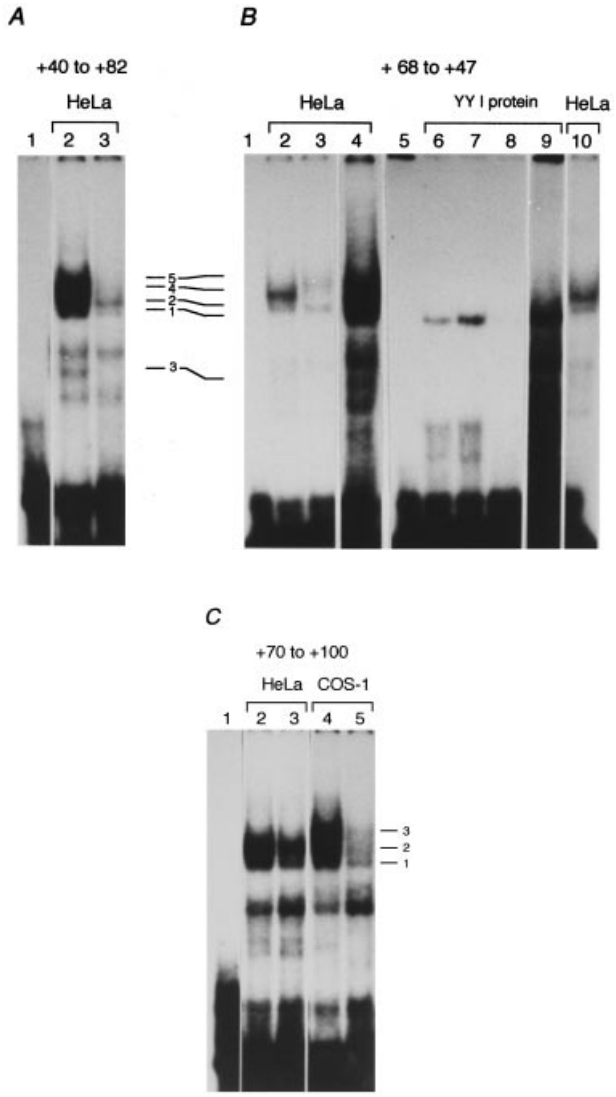


Figure 7 Analysis of intron 1 binding factors in COS-1 and HeLa nuclear extracts by gel retardation assay

(**A**) ^{32}P-labelled fragments (nt 40–82) were incubated in the absence (lane 1) or the presence of HeLa cell extracts (lanes 2 and 3). Lane 3 shows the effect of competition with 50 ng of unlabelled oligonucleotide 40–82. (**B**) Labelled oligonucleotide 68–47, carrying a possible YY1 binding site, was incubated without protein extracts (lanes 1 and 5), with HeLa cell extracts (lanes 2–4 and 10) or with bacterially expressed YY1 protein (lanes 6–9). For the experiments presented in lanes 6 and 7, 400 ng and 800 ng of recombinant YY1 protein was used respectively. Lanes 4 and 9 show the result of an incubation of labelled oligonucleotide 40–82 with HeLa cell extracts and recombinant YY1 protein respectively. Lanes 3 and 8 show competition with unlabelled oligonucleotide 68–47. A complex between a protein from HeLa cell extracts and oligonucleotide 68–47 is shown in lane 10. A 250 ng portion of poly(dG-dC) was used as a non-specific competitor with the recombinant protein, and 3 μg of poly(dG-dC) was used with HeLa cell extracts. (**C**) ^{32}P-labelled fragment 70–100 was incubated in the absence (lane 1) or the presence of HeLa cell extracts (lanes 2 and 5) or COS-1 cell extracts (lanes 3 and 4). The formation of a labelled protein–DNA complex was challenged by the addition of 50 ng of unlabelled competitor oligonucleotide (lanes 3 and 5). Poly(dG-dC) was used as a non-specific competitor.

nuclear factors is due to the flanking sequences (Figure 7B, lanes 2 and 4). As expected, bacterially expressed recombinant YY1 protein bound to both fragments (Figure 7B, lanes 6, 7 and 9) [19]. It appears that the recombinant YY1 protein has a higher binding affinity with the nt 40–82 oligonucleotide than with fragment nt 68–47, indicating that sequences around the core binding site are probably stabilizing the binding. However, the complex between bacterially expressed YY1 and oligonucleotide nt 68–47 runs faster on native gels than does the major complex between HeLa nuclear extracts and oligonucleotide nt 68–47 (Figure 7B, lanes 6, 7 and 10). This difference in mobility could be due to a post-translational modification or to proteolytic degradation.

The third oligonucleotide derived from intron 1 sequences (nt 70–100) was also used in a gel retardation assay with HeLa or COS-1 cell extracts. Two bands (numbered 1 and 2 in Figure 7C) were common to extracts from both cell types. At least one major slow-migrating band (Figure 7C, band 3) was only detected with the COS-1 cell extracts.

DISCUSSION

In this paper we describe *cis*-acting sequence elements in both the promoter region and the first intron of the rpS6 gene that are important for its transcription. Nuclear factors from COS-1 and HeLa cells were found to bind to sequences in the promoter region and first intron of the rpS6 gene. Some differences were detected between the factor binding capacities of the two cell extracts. The rpS6 gene, like a number of other ribosomal protein genes, requires a limited amount of 5′ upstream sequence for efficient transcription. Although sequences upstream of nt -100 and -64 are required for maximal activity in both COS-1 and HeLa cells, their contribution to the transcription of the rpS6–CAT gene is less than 40 %. Deletion of sequences upstream of nt -43 reduced transcription by 90 % in COS-1 cells. However, in HeLa cells, downstream transcriptional regulatory elements in the intron-containing rpS6–CAT gene can compensate for the deletion of the sequences upstream of -43 to restore about 50 % of transcriptional activity.

Thus we detected three additional nuclear factors in HeLa extracts which bind to the sequence between nt -33 and $+10$. This region contains a sequence (nt -22 to -16) which shows marked identity with the so-called χ-element motif first identified in the mouse rpL32 and mouse rpL30 promoters [10,14]. One of the transcription factors which binds to the χ-element motif also displayed a strong affinity for the canonical TATA box motif [20]. Similarly, recombinant transcription factor IID binds to the χ-element of mouse rpL32 [10]. Although this evidence is intriguing, transcription of rpL30 only depended on the presence of the χ-element when the promoter strength was decreased by deleting binding sites for other transcription factors [14].

Ribosomal protein genes and several other housekeeping genes have a GC-rich content in the vicinity of their transcription start sites. Conceivably these regions are potential targets for transcriptional regulatory factors such as Sp1, AP2 (adaptor protein 2), E2F and USF (upstream stimulatory factor) [21,22]. The presence of Sp1 binding site(s) in the promoter region is a common feature of housekeeping genes, especially in the CG-rich regions in and around their 5′ ends. Two Sp1 binding motifs are found in the 5′ upstream region of human rpS6 (nt -89 to -94 and nt -384 to -379), and one consensus sequence is present in the first intron (nt 241–246) [23]. However, 5′ deletion analysis indicates that the two upstream Sp1 motifs contribute less than 20 % of the transcriptional activity in COS-1 cells. In contrast, deletion of the region containing both of these upstream Sp1 binding elements seems to increase rather than decrease the expression of the rpS6–CAT gene in HeLa cells, indicating the possible presence of an inhibitory factor in HeLa cells which may bind to this region.

In an earlier report, a region called box A was described in the

5′ upstream region of the mouse rpL7a gene which was conserved in the corresponding position in the chicken and human rpL7a genes and appears to be crucial for maximal expression of mouse rpL7a transcriptional activity [24]. High identity with the rpL7a box A motif can be found in the region from nt −44 to nt −54 of the human rpS6 gene. In the band-shift analysis, at least three nuclear factors bound to regions between nt −64 and nt −33 containing this box A sequence.

The rpS6 5′ region also contains four possible binding sites for the ETS transcription factor family. All the ETS binding motifs are almost perfect fits for the proposed consensus sequences for the GABP (GA binding protein) factor, a member of the ETS DNA-binding protein family [25]. Using an oligonucleotide containing nt −19 to nt +1, including the potential GABP binding site, two complexes were formed with nuclear extracts from HeLa cells, while only the slower-migrating complex was formed with COS-1 cell extracts. A further GABP consensus motif can be found in the first exon (nt 6–12). Band-shift analysis of this region detected two retarded complexes with nuclear cell extracts from both cell types using oligonucleotides representing the entire exon 1 sequence. A *cis*-acting element called β has been described in the two mouse ribosomal protein genes rpL30 and rpL32 which exhibits a strong affinity for GABP binding proteins [26].

The presence of sequences from the first rpS6 intron results in a dramatic increase in gene expression. The effect of the rpS6 intron sequences on gene expression appears to be both context- and promoter-specific. A similar promoter-specific effect of intron sequences has been reported for immunoglobulin gene transcription, for which the intron-dependence is lost when the genes are regulated from a heterologous promoter [27]. Other genes, including those encoding the simian virus 40 T-antigen, mouse dihydrofolate reductase, purine nucleoside phosphorylase and manganese superoxide dismutase, also require intron sequences for efficient expression ([27,29,30] and refs. therein). Although other examples of ribosomal protein genes that require the presence of intron sequences for efficient expression have also been reported, efficient expression of mouse rpL7a appears to be independent of intron sequences [12,24,28].

The results presented here indicate that the intron-dependent increase in rpS6–CAT expression requires correct processing of the intron sequences. The replacement of rpS6 intron 1 by intron 2, which was in this context correctly spliced, was not sufficient to result in high expression. This suggests that the increase in gene expression may depend to some degree on the presence of transcription regulatory elements in the first intron. Deletion analysis of the intron 1 sequences revealed that sequence elements mainly localized in the 5′ region of intron 1 are the most important for expression. However, our observation that deletion of the middle part of intron 1 resulted in a partial loss of CAT activity supports the notion that, besides the exon 1–intron 1 boundary, other intronic regulatory elements may be important for the intron-dependent effect on rpS6 expression. In a similar fashion, sequence elements from the first intron have been shown to be important for rpL32 gene expression [28].

The intron 1 sequence of the human rpS6 gene contains a consensus binding site for the transcription factor YY1 (also called δ, nuclear factor-E1 and UCRBP) first identified in the Adeno-associated virus promoter. Putative YY1 binding motifs have been found in a number of ribosomal protein genes in both orientations of transcription. Deletion of the two YY1 binding sites located in the first exon and in the first intron of mouse rpL32 reduced transcription by about 90% [31–33]. Therefore one of the multiple *trans*-acting factors interacting with the 5′-terminal intron 1 sequence and detected in both cell extracts appears to represent the YY1 transcription factor. Besides binding sites that are often found within regulatory domains of housekeeping genes, the proximal intron 1 region contains binding sites for nuclear factors implicated in cell-specific transcription control. The sequence comprising nt 75–88 represents a binding motif for the liver-specific transcription factor C/EBP (CCAAT/enhancer binding protein), and nt 77–89 contain a possible binding motif which is recognized by the lymphocyte-restricted *Ikaros*-gene-encoded zinc finger DNA binding factors [34,35].

As might be expected for a housekeeping gene promoter, a high level of rpS6 gene transcription requires only a small region comprising exon 1, a very limited amount of flanking 5′ sequences and adjacent intron sequences. However, the reason for the multiplicity of band shifts and the differences (which may be cell-type-specific and/or species-related) between COS-1 and HeLa cell extracts obtained with most of the fragments tested remains unclear. Both COS-1 cells and HeLa cells are fast proliferating; therefore, if there is a common set of transcription factors that ensure high transcription levels, it should be present in both nuclear extracts. The differences may be due to the fact that the majority of the factors recognize the rpS6 sequences with low affinity and are in their function interchangeable, while binding of only a few obligatory factors is needed for high transcriptional activity. Alternatively, some factors may associate via protein–protein interactions, and the recruited partners, which may otherwise be involved in highly specific control of gene expression and vary from cell type to cell type, may serve in such a complex to ensure constitutive expression of housekeeping genes in a wide range of cells.

We thank Dr. Mike Fried for his support in the course of this work and in the preparation of the manuscript, and K. Gaston (Department of Biochemistry, University of Bristol, U.K.) for the recombinant protein YY1. We are grateful to Clive Dickson and Nadesan Gajendran for critical reading of the manuscript. This work was partly supported by a senior research fellowship to M.A. from the German Research Society (DFG)

REFERENCES

1 Amaldi, F., Bozzoni, I., Beccari, E. and Pierandrei-Amaldi, P. (1989) Trends Biochem. Sci. **14**, 175–178
2 Feo, S., Davies, B. and Fried, M. (1992) Genomics **13**, 201–207
3 Mager, W. H. (1988) Biochim. Biophys. Acta **949**, 1–15
4 Agrawal, M. G. and Bowman, L. H. (1987) J. Biol. Chem. **262**, 4868–4875
5 Curcic, D., Glibetic, M., Larson, D. E. and Sells, B. H. (1997) J. Cell. Biochem. **65**, 287–307
6 Loreni, F., Rubert, I., Bozzoni, I., Pierandrei-Amaldi, P. and Amaldi, F. (1985) EMBO J. **4**, 3483–3488
7 Beccari, E. and Mazzetti, P. (1987) Nucleic Acids Res. **15**, 1870–1872
8 Mariottini, P., Bagni, C., Francesconi, A., Cecconi, F., Serra, M. J., Chen, Q. M., Loreni, F., Annesi, F. and Amaldi, F. (1993) Gene **132**, 255–269
9 Chung, S. and Perry, R. P. (1993) Nucleic Acids Res. **21**, 3301–3308
10 Genuario, R. R., Kelley, D. E. and Perry, R. P. (1993) Gene Expression **3**, 279–288
11 Atchison, M. L., Meyuhas, O. and Perry, R. P. (1989) Mol. Cell. Biol. **9**, 2067–2074
12 Hariharan, N., Kelley, D. and Perry, R. P. (1989) Genes Dev. **3**, 1789–1800
13 Hariharan, N. and Perry, R. P. (1989) Nucleic Acids Res. **17**, 5323–5337
14 Safrany, G. and Perry, R. P. (1995) Eur. J. Biochem. **230**, 1066–1072
15 Kozma, S. C. and Thomas, G. (1992) Rev. Physiol. Biochem. Pharmacol. **119**, 123–155
16 Antoine, M. and Fried, M. (1992) Hum. Mol. Genet. **1**, 565–570
17 Smith, R., Peters, G. and Dickson, C. (1988) EMBO J. **7**, 1013–1022
18 Wildeman, A. G., Sassone-Corsi, P., Grundstrom, T., Zenke, M. and Chambon, P. (1984) EMBO J. **3**, 3129–3133
19 Gaston, K. and Fried, M. (1994) FEBS Lett. **347**, 289–294
20 Yoganathan, T., Horikoshi, M., Roeder, R. G. and Sells, B. H. (1993) FEBS Lett. **326**, 163–166
21 Stapleto, G., Somma, M. P. and Laria, P. (1993) Nucleic Acids Res. **21**, 2465–2471
22 Overman, P. F., Rhoads, P. D., Tasheva, E. S., Pyle, M. M. and Roufa, D. J. (1993) Somat. Cell Mol. Genet. **19**, 347–362

23 Kodonga, J. T., Jones, K. A. and Tjian, R. (1986) Trends Biochem. Sci. **11**, 20–23
24 Colombo, P. and Fried, M. (1992) Nucleic Acids Res. **20**, 3367–3373
25 Janknecht, R. and Nordheim, A. (1993) Biochim. Biophys. Acta **1155**, 346–356
26 Genuario, R. R. and Perry, R. P. (1996) J. Biol. Chem. **271**, 4388–4395
27 Banerji, J., Olson, L. and Schaffner, W. (1983) Cell **33**, 729–740
28 Chung, S. and Perry, R. P. (1989) Mol. Cell. Biol. **5**, 2075–2082
29 Jonsson, J. J., Foresman, M. D., Wilson, N. and McIvor, R. S. (1992) Nucleic Acids Res. **20**, 3191–3198
30 Jones, P. L., Ping, D. and Boss, J. M. (1997) Mol. Cell. Biol. **17**, 6970–6981
31 Hahn, S. (1992) Curr. Biol. **2**, 152–154
32 Hariharan, N., Kelley, D. E. and Perry, R. P. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 9799–9803
33 Chung, S. and Perry, R. P. (1993) Nucleic Acids Res. **21**, 3301–3308
34 Lopez-Rodriguez, C., Botella, L. and Corbi, A. L. (1997) J. Biol. Chem. **272**, 29120–29126
35 Molnar, A. and Georgopoulos, K. (1994) Mol. Cell. Biol. **14**, 8292–8303

Received 2 March 1998/21 August 1998; accepted 24 September 1998