Transcriptional Regulation of the Small Nuclear Ribonucleoprotein E Protein Gene

IDENTIFICATION OF CIS-ACTING SEQUENCES WITH HOMOLOGY TO GENES ENCODING RIBOSOMAL PROTEINS

(Received for publication, April 15, 1991)

Michael P. Fautsch† and Eric D. Wieben
From the Mayo Clinic/Foundation, Rochester, Minnesota 55905

The 5'-upstream region of the small nuclear ribonucleoprotein E protein (snRNP E) gene has multiple sequence similarities to elements found to be integral in the transcriptional regulation of some mouse ribosomal protein genes (G+C block, CTTCCG motifs) as well as U1 snRNA genes (U1 "B," U1 "D," and SPH enhancers). Furthermore, the immediate upstream region of the snRNP E protein gene lacks typical TATA and CAAT sequence motifs but contains one copy of a GC box characteristic of a housekeeping promoter. By transfection of constructs containing various amounts of the 5'-upstream region of the snRNP E protein gene linked to the bacterial gene for chloramphenicol acetyltransferase (CAT), we determined that the 5' boundary of sequence needed for transcription was within the first 153 nucleotides upstream of the transcription start site. Further deletions to within 51 base pairs reduced CAT expression by 2.5-fold. Mutational analysis within this 51-base pair region revealed that direct repeats of the hexamer CTTCCG were essential for CAT expression. Gel shift assays and DNase I footprinting experiments confirmed this conclusion by demonstrating specific binding of proteins to the CTTCCG motifs. This suggests that the direct repeats of the CTTCCG hexamer sequence play a key role in transcriptional regulation of the snRNP E protein gene.

Over the past decade considerable progress has been made in understanding the mechanism by which transcription is initiated by RNA polymerase II. In general, genes that are cell specific and expressed at high levels contain a promoter that encompasses a region near the transcription start site. This promoter region usually contains a match to the consensus sequence TATAAT (for review, see Breathnach and Chambon, 1981) which is involved in setting the transcription initiation site. Included in this promoter region are generally more distal elements that help play a role in controlling the frequency of initiation (CCAAT box, GC box). Recently, a variety of genes have been isolated which express proteins that function in a housekeeping or cellular growth control manner (Reynolds et al., 1984; Singer-Sam et al., 1984; Valebro, 1985; Ishii et al., 1985; Pfeffer et al., 1986; Melton et al., 1986; Farnham and Schimke, 1986; Hoffman et al., 1987; Boyer et al., 1988; Hong et al., 1989). All of these genes contain a transcription regulatory region that differs from the typical RNA polymerase II promoter region. Each has a high G+C content, most have multiple transcription initiation sites, and many do not contain the consensus sequence TATAAT. An interesting characteristic is that most housekeeping genes contain one or more copies of the GC box (5'-GCGGG-3') which may represent binding site(s) for the transcription factor Sp1 (Dyan, 1986).

The small nuclear ribonucleoprotein (snRNPs) E protein gene is another example of a constitutively acting gene. The snRNP E protein is a member of a core complex of proteins that, in association with the U family of small nuclear RNAs (snRNAs) and other associated snRNP proteins, is involved in the removal of intron sequences from primary transcripts (Maniatis and Reed, 1987; Sharp, 1987). The E protein multigene family consists of eight pseudogenes and one expressed gene that contains five exons (Stanford et al., 1986a, 1988b, 1991). This gene spans approximately 9 kilobases and encodes a basic protein that consists of 92 amino acids. Primer extension analysis in HeLa cells has identified two major transcription start sites (denoted as +1 and −3) and at least one minor transcription start site (Stanford et al., 1988b).

The 5'-flanking region of the snRNP E protein gene contains many sequence similarities to regulatory elements proposed to be involved in transcriptional regulation of some mouse ribosomal protein genes as well as U1 snRNA genes (Fig. 1). There is a G+C-rich region between −39 and −31 and two tandem repeats of the sequence motif CTTCCG which are found between −19 and −8. Two other copies of this hexamer sequence, both 5 out of 6-base pair matches, occur at −139 and at the main transcription start site. An inverted copy of the CTTCCG hexamer sequence is also found downstream of the main transcription start sites at positions +5 to +10. Previous studies have shown that at least one mouse ribosomal protein gene (L30) contains three copies of the sequence CTTCCG: two as upstream direct repeats and a third copy (5 out of 6-base pair match) spanning the transcription start site at positions −3 to +2. By using gel retention assays, Harihanan et al. (1989) were able to show in the ribosomal protein L30 gene that protein/DNA interactions

† The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04070.

‡ To whom correspondence should be sent: Mayo Clinic/Foundation, 200 1st St. S.W., Rochester, MN 55905. Tel.: 507-284-2299; Fax: 507-284-2053.

† The abbreviations used are: snRNP, small nuclear ribonucleoprotein; snRNA, small nuclear RNA; CAT, chloramphenicol acetyltransferase; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; CRE, cAMP-responsive element; TPA, 12-O-tetradecanoylphorbol-13-acetate; TRE, TPA-responsive element.
occurred within the region containing direct repeats of the CTTCCG sequence. Another mouse ribosomal protein gene, L32, also contains an inverted copy of the hexamer sequence CTTCCG from positions −76 to −71 (Atchison et al., 1989). Competition experiments for protein binding between these two DNA fragments found upstream of the L30 and the L32 genes suggest that a similar factor is involved in their binding (Harihanan et al., 1989).

Additional sequence similarities exist between transcriptional regulatory elements of the U1 snRNA gene and the snRNP E protein gene. Between −31 and −24 of the E protein gene is a 7 out of 8-base pair match to the sequence flanking the U1 "B" element (Murphy et al., 1987). This "B" element has been shown to be the sequence that fixes the start site of transcription in the human U1 snRNA gene. Strong homology is also found with two other elements that have been found to be important in regulating U1 snRNA transcription: the U1 "D" element and the SPH enhancer. The U1 "D" element is an enhancer element found upstream of the human U1 snRNA gene (Murphy et al., 1987). The snRNP E protein gene contains an 11 out of 12-base pair match to a sequence flanking this element at positions −135 to −122. The SPH enhancer sequence, which is found upstream of the chicken U1 snRNA gene, has a 10 out of 11-base pair match to the E protein gene sequence at nucleotides −99 to −89 (Roebuck et al., 1987, 1990).

Along with sequence similarities to transcriptional regulatory elements of the mouse ribosomal protein genes and the U1 snRNA genes, the snRNP E protein gene also contains two copies of a GC box. These sequences begin at positions −868 and −47. The immediate upstream region of the snRNP E protein gene does not contain a TATA box. The first occurrence of this sequence is at position −198. It is also interesting to note that 54 nucleotides upstream of this TATA box is an inverted CCAAT box (ATTGCC). Although these are found upstream, they are far from the usual position, which is generally 15–35 nucleotides for the TATA box and around 100 nucleotides from the transcription start site for the CCAAT box.

The existence of these sequence similarities raises the possibility that there may be some common regulatory mechanisms involved in modulating the expression of this group of genes. To determine if any of these similarities is important for the transcriptional regulation of the snRNP E protein gene, we present here a detailed functional characterization of its immediate 5′-upstream region. We have narrowed down the transcription regulatory region of this gene in human kidney 293 cells to within 153 nucleotides of the transcription start site and have identified several cis-acting elements, including the CTTCCG direct repeats, that are involved in the transcriptional regulation of the snRNP E protein gene.

**Materials and Methods**

Primer Extension Analysis—Sites of transcription initiation were determined by the procedure of Ebholtz et al. (1986) with the following modifications. Total RNA was isolated from human kidney 293 cells by homogenizing in guanidinium thiocyanate and pelleting the RNA through a cesium chloride gradient. 32P-end-labeled oligonucleotide DR5-3 (5′-GCTGCACCAAATTTTACACATAG-3′) was used as primer (Stanford et al., 1988b). 100 μg of total RNA was used per reaction. 250 units of Moloney murine leukemia virus reverse transcriptase were used in the extension reaction which was carried out at 37 °C for 45 min. After incubation, reactions were extracted with phenol, phenol/chloroform, and chloroform, precipitated with 2 M sodium acetate and ethanol, and air dried. The RNA was then run on a 10% denaturing polyacrylamide gel.

Plasmid Deletion Constructions—A genomic fragment of the E protein gene containing 2.6 kilobases of 5′-untranscribed sequence and 120 base pairs of transcribed sequence was cloned into Bluescribe(−) S1 nuclease digestion (200 units at 37 °C for 10 min) of the 3′ end of this insert followed by ligatation with HindIII linkers produced the clone S1H3-5, which contained all the upstream sequence plus 13 base pairs downstream of the transcription start site. Bal-31 exonuclease was added at a concentration of 20 units/ml at S1H3-5 that had been cleaved with the restriction enzyme KpnI. The sample was incubated at 37 °C, and aliquots were taken out 30-, 60-, and 90-s intervals. The DNA was isolated, and a fill-in reaction (Maniatis et al., 1982) was performed using the Klenow fragment of DNA polymerase I to obtain blunt-ended DNA. This DNA was ligated to HindIII linkers and cleaved with HindIII restriction enzyme to expose a 5′ overhang. These fragments were then digested to form circular plasmids and then transformed into competent HB101 cells. Clones were picked, analyzed for insert length, and appropriate length inserts were cloned into the HindIII site of the reporter plasmid pSV0-CAT (Gorman, 1985), which contains the bacterial gene for chloramphenicol acetyl transferase (CAT). CAT clones were sequenced by the methods of Sanger et al. (1977) for orientation and base composition.

Transfections, CAT and β-Gal Assays—For the transfection of CAT constructs into human kidney 293 cells, the procedure of Gorman et al. (1982) was followed with minor modifications. Kidney 293 cells (0.5–0.6 × 106) were seeded onto duplicate 100-mm Petri dishes and grown in Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum. Cesium chloride-purified plasmid DNA (10 μg) of each CAT construct was coprecipitated with 5 μg of control plasmid pH110 (containing the gene for β-galactosidase). The precipitate was added to cells and incubated for 10 min at 37 °C. Next, Dulbecco’s modified Eagle’s medium was added, and the cells were incubated at 37 °C for 24 h. After incubation, a 10% dimethyl sulfoxide shock was performed for 3 min at room temperature, and the cells were washed with Hank’s balanced saline to remove any residual dimethyl sulfoxide. Dulbecco’s modified Eagle’s medium was added, and the cells were incubated for an additional 49–72 h at 37 °C. Cells were rehydrated from the Petri dishes, and a cell lysate was made by freeze-thaw methods (5 min on dry ice and 5 min at 37 °C, repeated four
times). The protein concentration of each individual sample was determined by using the Bio-Rad protein concentration assay.

Chloramphenicol acetyl transferase assay reactions contained the following components: 70 μl of 1 M Tris-HCl, pH 7.8; 25 μg of human kidney 293 cellular extract; 20 μl of 40 mM acetyl coenzyme A; 1.5 μg of chloramphenicol (Du Pont-New England Nuclear, NEC-408); and water to a final volume of 150 μl. Samples were incubated at 37 °C for 1 h and the reactions stopped by the addition of 1 ml of ethyl acetate. The samples were freeze-dried, resuspended in 30 μl of ethyl acetate, and spotted on thin layer chromatography (TLC) plates in the presence of chloroform and methanol (95:5). CAT activity was quantified using the Ambis radioanalytic imaging system (model 1125). Each plasmid construct was tested in duplicate in at least six independent experiments using different DNA preparations. The CAT activity of kidney 293 cell extracts prepared from cells transfected with construct −153 to +13 was consistently greater than the positive control (pSV2-CAT) and routinely 15-fold higher than pSV0-CAT.

To standardize transfection efficiency, a β-galactosidase assay was performed on each sample by using 25–100 μg of human kidney 293 cellular extract. Incubation with o-nitrophenyl-β-D-galactopyranoside at a concentration of 3 mg/ml produced a yellow color which was measured on a Beckman spectrophotometer at wavelength 420 (Steers et al., 1971).

Nuclear Extract Preparation and Gel Retardation Assays—Nuclear extracts were prepared from human kidney 293 cells by the method of Dignam et al. (1983). Gel retardation assays were performed by the method of Singh et al. (1986) with minor modifications. Suitable restriction fragments were end labeled with T4 polynucleotide kinase in the presence of [32P]ATP (ICN Pharmaceuticals 35020). Binding reactions (15 μl) contained approximately 20,000 cpm of probe fragment (<1 ng), 3 μg of poly(dI-dC), 13 mM HEPES, 13% glycerol, 67 mM potassium chloride, 0.1 mM EDTA, 0.3 mM dithiothreitol, and 1–7 μg of human kidney 293 nuclear extract. After incubation for 30 min at 25 °C, protein–DNA complexes were separated from free DNA on nondenaturing 5% polyacrylamide gels made up in 50 mM Tris-HCl, 50 mM boric acid, 2 mM EDTA and run in the same buffer at 15 V/cm at 25 °C. Gels were dried down and exposed to Kodak diagnostic film XAR (165-1512).

DNase I Footprinting Experiment—DNase I footprinting analysis was performed as described previously by Galas and Schmitz (1978) with the following modifications. Binding reactions were carried out as described above except that the DNA fragment (−153 to +57) was 5’ end labeled at only one end with T4 polynucleotide kinase. After binding reactions, 1 unit of RQ1 DNase (Promega) was added to the reaction, and the final concentration of MgCl₂ was adjusted to 3 mM. Samples were incubated at room temperature for either 1 or 2 min and the reactions stopped by adding 25 μl of 0.1 M EDTA, 0.1% sodium dodecyl sulfate. Samples were extracted with phenol, phenolchloroform, and chloroform and ethanol precipitated. After centrifugation, the pellet was dissolved in formamide, boiled for 2 min, and loaded on a 10% polyacrylamide, 7 M urea gel. Gels were soaked in water for 10 min to remove urea, dried down, and autoradiographed.

Computer Analysis of DNA Sequences—The software packages of both DNA STAR (DNA STAR Inc., Madison, WI) and the University of Wisconsin Genetics Computer Group (UWGCC, Madison, WI) were used to analyze nucleotide homologies.

RESULTS

Analysis of Deletion Constructs—Before characterizing the 5’-upstream region, we first verified the transcription start sites for the snRNP E protein gene within human kidney 293 cells. By using primer extension analysis, two major and one minor transcription start site were identified (Fig. 2). These sites were similar to what was reported previously in HeLa cells (Stanford et al., 1988b). To determine if any of the sequence similarities to mouse ribosomal protein genes or U1 snRNA genes found within the 5’-flanking region of the snRNP E protein gene was important for transcriptional regulation, deletion constructs were made. Using Bal-31 exonuclease, deletion fragments were prepared from a genomic fragment of the E protein gene which contained 2.75 kilobase pairs of the 5’-upstream region. These deletion fragments were linked to the bacterial CAT gene and transfected into human kidney 293 cells (see “Materials and Methods”). Analysis of CAT activity in cells transfected with 10 μg of each construct revealed virtually no change in promoter activity when constructs containing 2.75 kilobase pairs of the 5’-upstream sequence or deletion constructs down to 153 base pairs of the 5’-upstream sequence were used (Fig. 3). Further deletion down to 80 base pairs decreased CAT activity in transfected cells by 50% when compared with the 153-base pair construct. The gradual increase in activity seen in Fig. 3 for constructs ranging from 2.75 kilobase pairs to 840 base pairs is probably caused by a difference in the molar amount of each deletion fragment used in the transfection procedure. Since equal amounts of DNA were added in the transfection procedure, the 2.75- and 1.5-kilobase constructs were underrepresented compared with the other constructs. When we performed a transfection experiment using equimolar amounts of the 2.75-kilobase pair construct and the 329-base pair construct, the CAT activity...
values from these cells were within 10% of each other (data not shown).

To delineate further the basal promoter activity of the immediate upstream region of the E protein gene, a double-stranded oligonucleotide was synthesized which contained 51 nucleotides of the 5'-upstream sequence (PC4; -51 to -46, Fig. 3). Cells transfected with this construct contained 2.5-fold less CAT activity than similar cells transfected with the 153-base pair construct. This suggested that there are positive regulatory elements between -153 and -51. However, the first 51 nucleotides of the 5'-flanking region were still sufficient to direct transcription.

**Mutational Analysis**—The first 51 nucleotides 5' of the transcription start site of the human E protein gene contain multiple sequence similarities to known transcriptional elements found in some mouse ribosomal protein genes. To determine which of these sequences might be involved in the transcriptional regulation of the E protein gene, selected mutations were made within these elements. Double-stranded oligonucleotides were synthesized which contained base pair changes at specific nucleotides within the GC box, G+C block, and the two direct repeats of the CTTCCG motifs (Fig. 4A). The annealed oligonucleotides were ligated to the CAT gene in the plasmid vector pSVO-CAT. Cellular extracts isolated from 293 cells transfected with equal molar concentrations of the m59 and m92 constructs which contained three mutations (each has a mutation within the GC box, G+C block, and the CTTCCG motif nearest the transcription start site) were found to have only 1/50 the CAT activity of similar extracts isolated from cells transfected with the PC4 construct (Fig. 4B). To determine which of these base pair changes within the m59 and m92 constructs were responsible for the reduction in CAT activity, further constructs were made. As seen in Fig. 4, single mutations within the GC box failed to reduce CAT expression (Spl-1, Spl-10, or Spl-12). Another construct was produced which replaced the GC box entirely (5'-CCGCC-3' with 5'-CATTGTT-3'). This also failed to reduce CAT expression significantly (data not shown). Cellular extracts isolated from 293 cells transfected with constructs GC-7, with a 1-base pair deletion within the G+C block, or GC-15, which has a 1-base pair transition from a C-G to a T-A, also fail to show a significant difference in CAT activity when compared with the PC4 construct. These results demonstrate that the GC box and the G+C block homologies are not required for promoter activity.

However, mutations within the two CTTCCG direct repeats immediately upstream from the transcription start site did result in reduced CAT activity in transfected 293 cells (Fig. 4). Construct 1CT-11 contains two mutations within the most 5'-CTTCCG motif whereas construct 2CT-16 contains two mutations within the CTTCCG motif closest to the transcription start site. 1CT-11 shows a 2-fold reduction in CAT expression whereas 2CT-16 exhibits approximately a 4-fold reduction when compared with the basal promoter construct PC4. These data suggest that the CTTCCG direct repeats are required for transcription of the E protein gene.

**Protein/DNA Interactions**—To investigate protein/DNA interactions that occur within the first 51 nucleotides upstream of the transcription start site, gel shift assays were performed with 32P end-labeled DNA fragments corresponding to PC4, 1CT-11, and 2CT-16 constructs. As can be seen in lane 4 of Fig. 5A, incubation of PC4 in the presence of nonspecific competitor poly(dI-dC) and nuclear extract prepared from human kidney 293 cells results in a gel shift that involves four bands (labeled I–IV). Incubation of construct 1CT-11 under the same conditions as wild-type PC4 fails to show shifted bands I and II (lane 6, Fig. 5A). Construct 2CT-16, which contains two mutations within the CTTCCG closest to the transcription start site, fails to yield bands I, II, and IV but does give a faint band at the position of band III (Fig. 5A; lane 5).

To confirm that the banding in the mobility shifts was caused by specific protein/DNA interactions at the CTTCCG motif, a competition experiment was set up using 50- or 100-fold molar excesses of unlabeled DNA fragments corresponding to PC4 (Fig. 5B; lanes 4 and 6) or 2CT-16 (Fig. 5B; lane 5). A 50-fold excess of the PC4 competitor results in a reduced intensity of bands I, II, and III (lane 4). Complete competition is seen with all four bands in the presence of a 100-fold molar excess of PC4 (lane 6). Competition with a 50-fold molar excess of 2CT-16, however, fails to compete with 32P end-labeled PC4 for bands I, III, and IV and only slightly with band II (lane 5). Taken along with results from Fig. 5A, this suggests that the mutations found within the CTTCCG sequence nearest the transcription start site specifically inhibit the binding of a protein or protein complex to the basal promoter fragment. To decipher which band(s) in the mobility shift corresponded to those interacting at the CTTCCG direct repeats, we made a double-stranded oligonucleotide (23CTT; see Fig. 5B) that spans the two CTTCCG direct repeats and ends just past the major transcription start site. The addition of unlabeled 23CTT in a 100- and 200-fold molar excess (Fig. 6; lanes 9 and 10) competed out all four bands but produced an additional band which can be seen in other reactions that contain competitor DNA (between bands III and IV).
FIG. 5. CTTCCG motifs are bound by protein. A, gel retention assay showing the amount of protein/DNA interactions found on fragments PC4, 2CT-16, and 1CT-11. All lanes contain approximately 1 ng of 32P end-labeled DNA and 5 μg of poly(dG)-(dC). Lanes 4–6 contain approximately 1 μg of human kidney 293 nuclear extract. Protein-DNA complexes are labeled I–IV. B, protein/DNA interactions on PC4 are specific. All lanes contain approximately 1 ng of 32P end-labeled PC4. Lanes 2–6 contain 7 μg of 293 cell nuclear extract. Lanes 3–6 contain 3 μg of poly(dI)-(dC). Lanes 4–6 also contain cold competitor DNA. The DNA fragment and the amount in molar excess are indicated above each lane. Note: Different preparations of kidney 293 nuclear extract were used for the experiments shown in A and B. This difference may explain the variance in binding intensities seen with band IV.

FIG. 6. Gel retention assay using specific cold competitor DNA that further identifies the CTTCCG direct repeats as regulatory elements. A, all lanes contain approximately 1 ng of construct PC4 that has been 32P end labeled, 7 μg of 293 cell nuclear extract, and 3 μg of poly(dI)-(dC). Cold competitor DNA was added as described above each line. 23CTT is a synthetically derived double-stranded oligonucleotide made by annealing complementary strands together at 37 °C before addition as a competitor. B, sequence of constructs used in gel retention assay are diagrammed except for 53111. 53111 is a DNA fragment spanning nucleotides −111 to −53 in the snRNP E protein gene’s upstream region. It was used as a nonspecific competitor. Constructs PC4, 2CT-16, and 1CT-11 contain 37 nucleotides 5′ and 33 nucleotides 3′ of additional sequence from pSVO-CAT. Asterisks represent base pair changes from the wild-type 51-base pair construct. Abbreviations of possible regulatory elements are described in Fig. 2.
new band may be caused by other protein/DNA interactions that take place in the −51 to −27 region of the PC4 construct. To investigate further the sequence specificity for the CTTCCG direct repeats, additional competition experiments were performed using 100- and 200-fold molar excesses of PC4, 2CT-16, 1CT-11, and 53111. 53111 is a double-stranded oligonucleotide that spans sequences −53 to −111 in the upstream region of the E protein gene. A 200-fold molar excess of 53111 failed to compete for any of the protein-DNA complexes formed on the ³²P end-labeled PC4 fragment (Fig. 6, lane 12). This again demonstrates the specificity with which proteins are bound to the PC4 fragment. In lanes 5–8 of Fig. 6, DNA fragments containing mutations within the CTTCCG direct repeats again failed to compete for the binding of protein(s) to the ³²P end-labeled wild-type fragment. An interesting observation is that 1CT-11, which by transfection experiments had only a 2-fold reduction in CAT activity compared with the wild-type PC4 construct, competed for protein binding, but the degree of competition was significantly less than that observed with the PC4 construct. This suggests that both CTTCCG repeats are capable of forming protein-DNA complexes. However, they may bind or help in the binding of protein(s) with different affinities (see “Discussion”).

To substantiate further the protein/DNA interactions occurring at the CTTCCG motifs, DNase I footprinting was used to identify sites of protein/DNA interactions. As can be seen in Fig. 7, lane 2, protein protection from DNase I digestion is seen from base pairs −30 to −8. This region contains the two direct repeats of the CTTCCG motif. A second region spanning the transcription start site is also protected from DNase I digestion (−5 to +5).

**DISCUSSION**

We have begun analysis of the transcriptional regulation of the snRNP E protein gene. The 5′-upstream region of this gene contains a housekeeping-type promoter with multiple sequence similarities to elements found to be essential in the transcriptional regulation of some mouse ribosomal protein genes as well as U1 snRNA genes. Deletion analysis of up to 2.75 kilobase pairs of the 5′-flanking sequence suggests that all elements essential for transcriptional activity in human kidney 293 cells reside within the first 153 base pairs upstream of the transcription start site. This indicates that the potential TATA box, the inverted CCAAT box, and the GC box found further upstream are not required for the function of this promoter.

Further deletions demonstrated that the basal promoter is found within the first 51 nucleotides. This short sequence yields 30% of the maximum CAT activity attained with longer constructs. Within this region are sequence similarities to a GC box, U1 “B” sequence, the G+C rich region, and the CTTCCG motifs. Mutations within the GC box or the G+C rich region failed to produce any change in CAT activity when compared with the PC4 construct. This suggests that the expression of the snRNP E protein gene does not require this GC box or the other upstream GC box for regulating its transcription.

Results from mutational analysis and subsequent gel mobility shift assays on the CTTCCG direct repeats identify them as important for the transcription of the E protein gene. Although mutations within the CTTCCG nearest the transcription start site affect CAT expression to a greater extent than the other CTTCCG (4-fold versus 2-fold reduction when compared with the wild-type PC4 construct), mobility shift analysis suggests that both elements are capable of participating in protein/DNA interactions. Because different base pairs are mutated within each CTTCCG (Fig. 3), more definitive mutational analysis within the two repeats will need to be done to see if the protein(s) contact points identify particular base pairs with a higher affinity than others. A difference in protein affinity for particular base pairs within the CTTCCG direct repeats could account for the discrepancy that we see in CAT expression as well as gel mobility shift assays involving competition experiments seen between the two CTTCCG sequences. It is apparent, however, that both CTTCCG motifs are important in regulating transcription of the snRNP E protein gene. Another explanation for the difference seen between the two CTTCCG motifs in transfection experiments and mobility shift assays could be the importance of the flanking sequence. Rhoads et al. (1986) identified a short sequence motif in the 5′-upstream region of the human ribosomal protein S14 gene which is found in related versions in other ribosomal protein genes isolated from yeast and mouse (Fig. 8A). The E protein gene also contains two overlapping sequences with similarity to this motif at positions −21 to −6 (Fig. 8B). As can be seen in this figure, the core of each of these two sequences is the CTTCCG hexamer sequence. As mentioned in the Introduction, Hariharan et al. (1989) identified a factor(s) that binds to nucleotides −75 to −52 in the mouse ribosomal protein L30 gene. Within this 24-base pair stretch are two direct repeats of the CTTCCG motif. It is interesting to note that the short sequence motifs identified by Rhoads et al. (1986) found in the ribosomal protein S14 gene’s 5′-upstream region are also present in this 24-base pair region of the ribosomal protein L30 gene (Fig. 8B). The significance of this sequence and the factor(s) bound by it are unknown. However, its presence within multiple ribosomal protein genes as well as its identification within the human snRNP E protein gene make it an attractive candidate for transcriptional regulation.

Another area of interest is the region between −153 and −51. Transfection of equal molar amounts of the −153 and −51 constructs showed a reduction in CAT activity in transfected human kidney 293 cells by 2.5-fold. This suggests that at least one transcriptional regulatory element exists between −153 and −51. The deletion to −51 removed two potential enhancer elements identified by computer analysis (the U1 “D” element and the U1 SPH enhancer element). Two other
Transcriptional Regulation of the snRNP E Protein Gene

A

| Gene     | Species | Position | Sequence       |
|----------|---------|----------|----------------|
| Ribosomal proteins: | | | |
| L30      | Mouse   | -65      | TCTCCCGCTT     |
|          |         | -59      | GCCCTCGGCCT    |
| L32      | Mouse   | -70      | *ACTTCCGGCT    |
| S14      | Human   | -69      | TCTTCCGGAA     |
|          |         | -29      | ACTTCCGGAA     |
| snRNP proteins: | | | |
| E        | Human   | -20      | ACTTCCGGCTT    |
|          |         | -14      | GCCCTCGGGTT    |
|          |         | +11      | *ACTTCCGGAA    |

B

Ribosomal protein L30

\[
\begin{array}{c}
\text{GACCGGCTTCTCCCGCGGTCCCGTCCCGTCCCGGCTCGGGCTCCGGAAGTTGCTTCGAGG} \\
\text{GACCGGCTTCTCCCGCGGTCCCGTCCCGTCCCGGCTCGGGCTCCGGAAGTTGCTTCGAGG} \\
\text{GACCGGCTTCTCCCGCGGTCCCGTCCCGTCCCGGCTCGGGCTCCGGAAGTTGCTTCGAGG} \\
\end{array}
\]

snRNP E protein

\[
\begin{array}{c}
\text{TGACCTCACTCCTCCTCGGTCTCCCTTATTACCTCGGGAAGTTTGCTTCGAGG} \\
\text{TGACCTCACTCCTCCTCGGTCTCCCTTATTACCTCGGGAAGTTTGCTTCGAGG} \\
\text{TGACCTCACTCCTCCTCGGTCTCCCTTATTACCTCGGGAAGTTTGCTTCGAGG} \\
\end{array}
\]

Table 1

Sequence similarities of TRE/CRE elements

Sequences were identified from the following references: CRE (Montminy et al., 1987); c-jun TRE (Karin, 1991); consensus TRE (Wingender, 1990).

| CRE (cAMP) | AP-1 (Jun) | E protein | AP-1 or TRE |
|------------|------------|-----------|-------------|
| TGACGCTCA  | GTGACCATCAT| GTGACATCAT | TGACATCAT    |
| TCAG_6     | TCAG_6     | TCAG_6    | TCAG_6      |

features within the -153 to -51 region are worth mentioning. The first is the presence of a 12 out of 13-base pair match to a sequence found upstream of the La antigen gene (−109 to −97) (Chambers et al., 1988). Although no function to this region has been reported in the La antigen 5′-flanking region, the high sequence similarity and position conservation within the two upstream regions raises the possibility that this region has significance.

The second potentially interesting characteristic is the occurrence of a potential TPA-responsive element (TRE) or cAMP-responsive element (CRE) at positions −81 to −72 (Table I). The sequence found in the snRNP E protein gene differs from the TRE element found in the human c-jun promoter by one nucleotide. If this same nucleotide is switched to a G, the element becomes cAMP responsive. Angel et al. (1988) have shown that the TRE site in the human c-jun gene is recognized by a high affinity AP-1 complex. They have also suggested that this AP-1 complex does not appear to contain c-fos (Karin, 1991). This implies that different forms of the AP-1 complex may exist which can bind variants of the TRE element. Further examination of this region as well as the other sequence similarities found within the −153 to −51 region will lead to a better understanding of their possible role in the transcriptional control of the E protein gene.

With the previous identification of the CTTCCG motifs involved in the transcriptional regulation of some ribosomal protein genes and now their regulatory importance determined upstream of the snRNP E protein gene, it is interesting to speculate that a possible mechanism for controlling transcription exists between some components of the spliceosome and the ribosome. Etzerodt et al. (1988) suggested that the developmental patterns of the mRNAs of the snRNP E protein gene and ribosomal protein genes are expressed in a similar fusion in Xenopus. They suggested the possibility that a common mechanism for controlling transcription of these genes may exist. Further experimentation and identification of other promoter regions of ribosomal protein genes as well as snRNP core proteins will help to determine if a relationship exists between the transcriptional regulation of the snRNP E protein gene and ribosomal protein genes.

Acknowledgments—We wish to thank Drs. Dana Kim Reed, Jane Toft, and April Chang-Miller for helpful comments on this manuscript. We also wish to thank Debbie Pearson for help in preparing this manuscript.

REFERENCES

Angel, P., Hattori, K., Smeal, T., and Karin, M. (1988) Cell 55, 875–885

Atchison, M. L., Meyuhas, O., and Perry, R. P. (1989) Mol. Cell. Biol. 9, 2067–2074

Bayer, T. G., Krug, J. R., and Maquat, L. E. (1989) J. Biol. Chem. 264, 5177–5187

Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–385

Chambers, J. C., Kenan, D., Martin, B. J., and Keene, J. D. (1988) J. Biol. Chem. 263, 18043–18051

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) Nucleic Acids Res. 11, 1475–1489

Dynan, W. S. (1986) Trends Genet. 2, 196–197

Elsholtz, H. P., Mangalam, H. J., Potter, E., Albert, V. R., Supowit, S., Evans, R. M., and Rosenfeld, M. G. (1986) Science 234, 1552–1557

Etzerodt, M., Vignoli, R., Cibert, G., Scherly, D., Mattaj, I. W., and Phillipson, L. (1988) EMBO J. 7, 4311–4321

Farinham, P. J., and Schimke, R. T. (1986) Mol. Cell. Biol. 6, 2392–2403

Galas, D. J., and Schmitz, A. (1978) Nucleic Acids Res. 5, 3157–3170

Gorman, C. (1985) in DNA Cloning: A Practical Approach (Glover, D. M., ed) Vol. 11, pp. 143–190, IRL Press, Washington, D.C.

Gorman, C. M., Moffat, L. F., and Howard, H. B. (1982) Mol. Cell. Biol. 2, 1044–1051

Harbharan, N., Kelley, D. E., and Perry, R. P. (1989) Genes & Dev. 3, 1789–1800

Hoffman, E. K., Trusko, S. P., Freeman, N., and George, D. L. (1987) Mol. Cell. Biol. 7, 2392–2396

Hong, F. D., Huang, H. S., To, H., Young, L. S., Oro, A., Bookstein, R., Lee, Y. H. P., and Lee, W. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 5502–5506

Ishii, S., Yu, Y. H., Stratton, R. H., Roe, B. A., Merlino, G. T., and Pasten, I. (1985) Science 229, 1378–1381

Karin, M. (1991) in Molecular Aspects of Cellular Regulation (Cohen, P., and Powikes, G., eds) Vol. 6, pp. 143–161, Elsevier, New York

Maniatis, T., and Reed, R. (1987) Nature 325, 673–678

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, pp. 113–114, Cold Spring Harbor
Transcriptional Regulation of the snRNP E Protein Gene

Laboratory, Cold Spring Harbor, NY
Melton, D. W., McEwan, C., McKie, A. B., and Reid, A. M. (1986) Cell 44, 519–528
Montminy, M. R., and Bilezikjian, L. M. (1987) Nature 328, 175–178
Murphy, J. T., Skuzeski, J. T., Lund, E., Steinberg, T. H., Burgess, R. R., and Dahlberg, J. E. (1987) J. Biol. Chem. 262, 1799–1803
Petel, P. I., Framson, P. E., Caskey, C. T., and Chinnault, A. C. (1986) Mol. Cell. Biol. 6, 393–403
Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L., and Luskey, K. L. (1984) Cell 38, 275–285
Rhoads, D., Dixin, A., and Roufa, D. J. (1986) Mol. Cell. Biol. 6, 2774–2783
Roebuck, K. A., Walker, R. J., and Stumph, W. E. (1987) Mol. Cell. Biol. 7, 4186–4193
Roebuck, K. A., Szeto, D. P., Green, K. P., Fan, Q. N., and Stumph, W. E. (1990) Mol. Cell. Biol. 10, 341–352
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
Sharp, P. A. (1987) Science 235, 766–771
Singer-Sam, H., Keith, D. H., Tani, K., Simmer, R. L., Shively, L., Lindsay, S., Yoshida, A., and Riggs, A. D. (1984) Gene (Amst.) 32, 409–417
Singh, H., Sen, R., Baltimore, D., and Sharp, P. A. (1986) Nature 319, 154–158
Stanford, D. R., Kehl, M., Perry, C. A., Holicky, E. L., Harvey, S. E., Rohleder, A. M., Rehder, K., Luhrmann, R., and Wieben, E. D. (1988a) Nucleic Acids Res. 16, 10593–10605
Stanford, D. R., Perry, C. A., Holicky, E. L., Rohleder, A. M., and Wieben, E. D. (1988b) J. Biol. Chem. 263, 17772–17779
Stanford, D. S., Holicky, E. L., Perry, C. A., Rehder, K., Harvey, S. E., Rohleder, A. M., and Wieben, E. D. (1991) J. DNA Sequence 1, 357–363
Steers, E., Cuatrecasas, P., and Pollard, H. P. (1971) J. Biol. Chem. 246, 196–200
Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M., Weeda, G., Berkvens, T. M., van der Voorn, L., van Ormondt, H., and van der Eb, A. J. (1985) EMBO J. 4, 437–443
Wingender, E. (1989) Crit. Rev. Eukaryotic Gene Expression 1, 11–48