Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements

Narayanan Hariharan, Dawn E. Kelley, and Robert P. Perry

Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111 USA

The promoters of the mouse ribosomal protein genes *rpL30*, *rpL32*, and *rpS16* are of equal strength, as indicated by *in vivo* measurements of polymerase loading and by their relative efficiency in driving the expression of a linked reporter gene. The equipotency of these promoters appears to derive from a remarkably similar architecture in which five or more elements are distributed over a 200-bp region that spans a polypyrimidine-embedded cap site. Three *trans*-acting factors are shared by the *rpL30* and *rpL32* promoters, one of which, δ , recognizes a common CNGCCATCT motif in the first (untranslated) exons. Site-specific mutagenesis demonstrated that δ -factor binding is critical for *rpL30* promoter function. The repeated occurrence of this novel promoter architecture among ribosomal protein genes with very different coding specificities is most readily explained by convergent evolution.

[Key Words: Promoters; convergent evolution; ribosomal proteins]

Received May 26, 1989; revised version accepted August 17, 1989.

The mammalian ribosome contains more than 70 different species of proteins that are coordinately synthesized during growth and development. In murine cells of diverse tissue origins, the balanced synthesis of the various ribosomal proteins is largely attributable to the fact that the respective ribosomal protein mRNAs are present in similar amounts and are translated with comparable efficiencies (Meyuhas and Perry 1980; Meyuhas et al. 1987). As will be demonstrated below, the uniformity in ribosomal protein mRNA content is paralleled by a uniformity in the rate of transcription of the corresponding ribosomal protein genes. Because ribosomal protein genes are not generally clustered in the mouse genome (Wiedemann et al. 1987), it seems reasonable to suppose that their relative transcription rates are mainly determined by the relative strengths of their promoters.

To explore the basis of ribosomal protein promoter strength, we decided to compare several ribosomal protein promoters with respect to their modular *cis*-acting elements and the proteins that specifically bind to these elements. In particular, we wished to ascertain whether different ribosomal protein promoters contain common elements and whether there is any discernible similarity in the spatial organization of the elements. For this purpose we selected three mouse ribosomal protein genes, *rpL30*, *rpL32*, and *rpS16*. These single-copy genes are located on different chromosomes (Wiedemann et al. 1987) and are evolutionarily unrelated, as judged from the complete dissimilarity of their sequences (Dudov and Perry 1984; Wiedemann and Perry 1984; Wagner and Perry 1985). In this paper, we present a detailed analysis

of the *rpL30* promoter and compare its properties with those of *rpL32* (Dudov and Perry 1986; Atchison et al. 1989; Chung and Perry 1989; Moura-Neto et al. 1989) and *rpS16* (Hariharan and Perry 1989).

Our findings reveal that all three ribosomal protein promoters have a similar architecture, which is distinguished from that of other well-characterized polymerase II promoters by having essential elements located downstream of a polypyrimidine-embedded cap site. Moreover, the resemblance between the *rpL30* and *rpL32* promoters is particularly striking. The arrangement of modular elements within these promoters is very similar, and three out of five correspondingly located modules interact with the same nuclear factor or with factors that have identical bindings specificities. These novel ribosomal protein promoters, which are essentially equipotent when used to drive the expression of a linked reporter gene, have most likely originated by convergent evolution.

Results

rpL30, *rpL32*, and *rpS16* genes are transcribed at equal rates

In previous studies of various cultured murine cell lines, namely, plasmacytoma, fibroblast, and lymphosarcoma cells, we observed that the steady-state content of mRNAs encoding L30, L32, and S16 was essentially the same (Meyuhas and Perry 1980; Meyuhas et al. 1987). To

determine whether this uniformity in mRNA content reflects a uniformity in the rate of transcription of the corresponding ribosomal protein genes, we measured the relative polymerase loading of rpL30, rpL32, and rpS16 by nuclear run-on experiments. For these measurements we used ribosomal protein gene fragments (Fig. 1A) that are known to be free of repetitive sequences (Dudov and Perry 1984; Wiedemann and Perry 1984; Wagner and Perry 1985) and nuclei prepared from two immunoglobulin-producing plasmacytoma cell lines, MPC11 and S194. For comparison, polymerase loading of the highly active κ immunoglobulin genes was also measured. The results of these experiments (Fig. 1B,C) demonstrated that each of the three ribosomal protein genes is equally loaded with polymerase, the observed variance of $\pm 7\%$ being within the limits of accuracy for this type of measurement. In MPC11 cells, the loading of the individual

ribosomal protein genes was $\sim 7\text{--}10\%$ that of the κ genes, a value commensurate with the relative abundance of ribosomal protein and κ mRNA in these cells (Meyuhas and Perry 1980). Furthermore, the transcription of the ribosomal protein and κ genes was equivalently inhibited by α -amanitin (Fig. 1B,D), verifying that ribosomal protein genes are indeed transcribed by RNA polymerase II (Lindell et al. 1970).

The uniform in vivo transcription rate of these unlinked ribosomal protein genes probably reflects a similarity in the strength of their promoters. To determine whether these ribosomal protein promoters have any common structural features that could account for this apparent similarity in promoter strength, we have defined the modular elements of the rpL30 promoter and have compared them with their rpL32 and rpS16 counterparts.

Figure 1. Relative polymerase loading of ribosomal protein genes. (A) Schematic diagram showing the exon–intron organization of the ribosomal protein genes and the DNA segments (thick bracketed lines) used for the nuclear run-on assays. (Boxes) Exons; (hatched portions) untranslated regions; ($\blacktriangledown$, $\triangledown$) start and stop codons, respectively; (thin lines) introns and flanking regions; (horizontal and vertical arrows) transcriptional start sites and poly(A) sites, respectively. Numbers give the sizes of each segment in base pairs. The total sizes of the transcription units and the mature mRNAs (parentheses) are given at the upper left of each gene. (B) Autoradiograms of ^{32}P -labeled nuclear RNA hybridized to DNA dots containing segments of the three ribosomal protein genes, the κ immunoglobulin gene, and a plasmid vector pBR322 (pBR). The κ gene segment, which was derived from the genomic clone pEC κ (Coleclough et al. 1981), encompasses ~ 5 kb of transcribed sequence. Because of the difference in effective probe size, the ratio of κ to ribosomal protein hybridization signals is 2–2.5-fold greater than the ratio of polymerase loading. The RNA was isolated from MPC11 plasmacytoma nuclei that were incubated for 10 min at 26°C with a [α - ^{32}P]UTP in the absence (–) or presence (+) of $1\ \mu\text{M}$ α -amanitin (α A). (Right) Numbers represent the dot intensities, corrected for background, divided by the relative length of hybridizing sequence and normalized to the average for the three ribosomal protein genes (– α A). (C) Relative polymerase loading of ribosomal protein genes. Data from four experiments (two with MPC11 nuclei and two with S194 nuclei) were analyzed as in B, and the average values and deviations plotted as a bar graph. (D) Percent residual ribosomal protein and κ transcription as a function of α -amanitin concentration. ($\circ$) κ ; ($\times$) rpL32; ($\bullet$) rpL30; ($\blacktriangle$) rpS16.

Functional analysis of the rpL30 promoter

In an initial set of experiments, a series of rpL30 genes containing progressively shorter lengths of 5'-flanking sequence, the complete internal portion of the gene and 260 bp of 3'-flanking sequence (a-e, Fig. 2) were transfected into monkey kidney (COS-7) cells and their transient expression assayed by S1 nuclease protection. The S1 probe was designed to yield a 61-nucleotide fragment when protected by a properly spliced rpL30 transcript that is initiated at the authentic rpL30 cap site (Fig. 3). The background that results from cross-reactivity of this probe with the L30 mRNA produced by endogenous COS cell genes is sufficiently low so that a reasonably sensitive assay of the relative activity of the transfected genes can be made. As a control for equivalent transfection efficiencies and RNA yields, we monitored the expression of a cotransfected intact rpS16 gene. As seen from the data of Figure 3A, rpL30 expression was reduced by 40% when sequences between -205 and -94 were deleted and by 80% when the deletion extended to -52. Further removal of the -51 to -31 region did not lower the expression additionally; but when the deletion was extended to -9, expression dropped to the background level observed with untransfected COS cells.

Recent studies of the rpL32 gene (Chung and Perry 1989) have revealed an important transcriptional regulatory element at the 5' end of its first intron. To investigate whether the first intron of rpL30 is also important for expression, we used a highly homologous processed pseudogene, L30-2 (Wiedemann and Perry 1984), to construct a recombinant that is essentially identical to the complete rpL30 gene (construct a, Fig. 2), except that it lacks the first intron (construct f, Fig. 2). When this re-

combinant gene was transfected into COS cells, there was no detectable expression (Fig. 3B), indicating that there is indeed a critical element within the first intron of rpL30.

The various *cis*-acting elements were further localized with constructs in which the rpL30 promoter region was linked to the chloramphenicol acetyltransferase (CAT) reporter gene. Because we did not have to be concerned about endogenous ribosomal protein gene activity with the CAT assay system, we chose to use mouse plasmacytoma cells (S194) for this series of transfection experiments. In one group of constructs (g-k, Fig. 2), the CAT gene was inserted upstream of the ATG initiation codon in exon II and the amount of 5'-flanking sequence was varied between -205 and -52. The results of these experiments, summarized in Figure 2, localized the most upstream element to the -135 to -94 region and an adjacent element to the -75 to -52 region. The relative decrease in expression caused by the successive removal of these elements (40% and 80%, respectively) was essentially the same as that observed in COS cells when the entire rpL30 gene was present. Thus, the relative promoter activity of the CAT constructs faithfully represents that of the intact rpL30 gene. Moreover, this promoter functions equivalently in murine and simian cells.

A second group of constructs (l-p, Fig. 2) contained 205 bp of 5'-flanking sequence and various amounts of internal sequence. Consistent with the result obtained with construct f, constructs l and m, which do not contain any intron 1 sequences, were expressed very poorly, i.e., at ~6-8% of the maximum level. In constructs n, o, and p, the CAT gene was joined to a 73-bp rpS16 segment, which supplies a functional 3' splice site; this

Figure 2. The various rpL30 constructs used in transfection experiments and their relative level of expression. The 5' end of the gene is shown with exons I and II as solid bars and the intron and flanking region as thin lines; scale is in base pairs. The CAT reporter gene is indicated by an open rectangle; an inserted 3' splice junction segment is indicated by a crosshatched tapered bar. (Right) The relative expression normalized to construct a (for a-f) and g (for g-p). Expression assays were as illustrated in Fig. 3. Values represent averages of at least two independent transfection experiments. The CAT activities were measured with an AMBIS Radioanalytic System and standardized to the relative amount of plasmid DNA in Hirt supernatant fractions of the transfected cells.

unit was then linked to rpL30 promoter fragments that extend 11, 39, and 54 bp into intron 1. The insertion of a functional 3'-splice site between an unpaired 5' splice site and the CAT gene is necessary to obtain maximum levels of CAT activity (see Materials and methods). The expression of constructs n, o, and p was 16%, 100%, and 90% of maximum, respectively (Figs. 3C and 2), indicating that the critical intronic element is located within the first 39 bp of intron 1. The same results were obtained when constructs g, l, m, and o were transfected into COS cells (data not shown).

The above results indicate that the rpL30 promoter contains at least four discrete elements in the region that extends from 135 bp upstream to 77 bp downstream of the cap site. Data presented in Figure 7 demonstrate the existence of an additional element within exon I. These functionally important elements, designated A–E for convenience, are diagramed as bracketed regions in Figure 4. When the overall structure of the rpL30 promoter was compared to that of the rpL32 and rpS16 promoters (Fig. 4), several similarities were noted, e.g., an array of five to six elements embedded in a region of ~200 bp, the critical involvement of elements located downstream of the cap site, and the stimulatory but nonessential role of the most upstream elements. These similar features may be hallmarks of a common ribosomal protein promoter architecture, a concept that is further examined by the factor-binding studies detailed below.

Location of binding sites for rpL30 transcription factors

The rpL30 promoter region was surveyed for factor-binding sites by the gel-mobility shift assay (Fried and Crothers 1981). A series of DNA fragments covering the –205 to +86 region of the rpL30 gene were assayed with nuclear extracts prepared from S194 plasmacytoma

cells. These assays localized binding sites to four regions of the rpL30 promoter: α , between –135 and –94; β , between –75 and –52; γ , between –31 and –9; and δ , between –4 and +49 (Fig. 5). The specificity of binding was verified by the differential effects of unlabeled rpL30 competitor fragments. For α , an excess of –135 to +77 fragment completely abolished binding, whereas an excess of –94 to +77 fragment had no effect (Fig. 5A, lanes 3 and 4). For β , there was effective competition with a –75 to +77 fragment and no competition with a –52 to +77 fragment (Fig. 5B, lanes 3 and 4). The reason for the complexity of the β -factor retardation pattern is presently unknown; it could be due to the existence of multiple forms of the factor or to interactions with other proteins. Similar competition experiments confirmed the specificity of binding to the γ and δ sites (Fig. 5C, lanes 3 and 4; Fig. 5D, lanes 3 and 4). The factors binding to the α , β , γ , and δ sites are clearly different from each other, as demonstrated by the appropriate competition experiments (Fig. 5A–C, lane 4).

Three of the regions in which no binding was detected by this assay, i.e., –205 to –135, –94 to –75, and –52 to –31, correspond to regions that do not contribute significantly to promoter activity, as judged by the functional analyses described above. However, the +49 to +86 region also tested negative by the gel-shift assay, although it encompasses the portion of intron 1 that was shown to be critical for efficient performance of the rpL30 promoter. Possible reasons for this apparent anomaly are considered in the Discussion.

Extent of factor sharing among the ribosomal protein promoters

The identification of nuclear factor-binding sites in comparable regions of the three ribosomal protein promoter regions (see Fig. 4) raised the question of whether

B', C, and D, respectively. The prime symbols designate elements that were demonstrated to be functionally important by transfection analyses but are not yet known to possess a factor-binding site.

Figure 4. Comparative organization of the rpL30, rpL32, and rpS16 promoters. The functionally important regions, localized by transfection experiments with appropriately modified constructs, are shown as bracketed segments, A–E, below each gene. The width of a segment, which reflects the particular constructs used in the experiment, may be greater than the boundary of the element contained within it. Dotted lines designate regions that do not appear to have functional importance. An uncertainty about the exact boundary of an element is indicated by a dashed vertical line. The segments containing specific binding sites for nuclear factors are shown as boxes above each gene. Boxes of the same pattern denote segments that compete for a common factor. Boxes marked with X indicate segments that exhibit complex patterns in gel mobility-shift analysis. The data for rpL32 and rpS16 are described in Atchison et al. (1989) and Hariharan and Perry (1989), respectively. To have a uniform nomenclature for presumably identical factors, the rpL32 factors previously designated as α and β are renamed β and γ , respectively. The rpS16 elements previously designated as C, D, and E have been renamed

any of the factors is shared among the different promoter regions. To investigate this possibility, we included selected fragments of rpL32 and rpS16 in the competition experiments of Figure 5 and also carried out reciprocal competition experiments with labeled rpL32 and rpS16 fragments. The L30- α -region binding was not affected by excess competitor fragments containing the functionally relevant 5'-flanking regions of rpL32 and rpS16 (Fig. 5A, lanes 5 and 6), indicating that the L30- α factor is not shared by the other two ribosomal protein genes. In contrast, β -region binding was efficiently competed out by an L32 -79 to +11 fragment, but not by L32 -68 to +11 or +16 to +77 fragments (Fig. 5B, lanes 5–7), indicating that the L30- and the L32- β factors have the same or very similar binding specificities. A reciprocal experiment, in which specific factor binding to a labeled fragment containing the L32- β site was competed out by excess fragment containing the L30- β site, confirmed the specificity of this competition assay and further indicated that the affinity of the β factor(s) for the rpL30 and rpL32 sites is not vastly different (data not shown). An rpS16 promoter fragment did not efficiently compete for either rpL30- or rpL32- β factor binding (Fig. 5B, lane 8 and data not shown).

The factor binding to the rpL30- γ region also binds to a similar region of the rpL32 gene. This is evidenced by effective cross-competition between the corresponding L30 and L32 fragments but not by fragments from other rpL30 regions or from rpS16 (Fig. 5C, lanes 4–7 and data not shown). The δ -region factor is also recognized by rpL32. Two rpL32 fragments efficiently competed for binding to the rpL30- δ region: one representing a similar segment of exon I sequences and another containing a segment of intron 1 sequences (Fig. 5D, lanes 6 and 7). No competition was observed with a more upstream

rpL32 fragment or with a fragment containing the important internal sequences of the rpS16 gene (Fig. 5D, lanes 5 and 8). In a reciprocal experiment, binding to the rpL32- δ region was effectively competed out by excess rpL30- δ -region fragment but not by fragments from rpS16 (data not shown). The rpL32 intron 1 fragment that effectively competed for the L30- δ factor (Fig. 5D, lane 7) contains the binding site designated as ϵ (Fig. 4), which was localized to the +53 to +77 region of rpL32 (Atchison et al. 1989). Binding to the L32- ϵ site is effectively competed out by fragments containing the δ -factor site from either rpL30 or rpL32 but not by rpL30 fragments containing the γ -factor site or the +50 to +79 region of intron 1 (data not shown). This suggests that δ and ϵ may be the same factor or factors with similar binding specificities.

The foregoing results indicate that three of the nuclear factors that bind to elements of the rpL30 promoter, β , γ , and δ , also bind to elements in comparable locations of the rpL32 promoter; the fourth factor, L30- α , is apparently not used by the rpL32 promoter. None of the rpL30 or rpL32 factors is recognized by the rpS16 promoter. This was evidenced by the consistent lack of competition of rpS16 fragments for rpL30 and rpL32 binding and by a similar lack of competition of L30 and L32 fragments for rpS16-factor binding (data not shown). Thus, despite the similarities in overall promoter architecture, the individual elements of the rpS16 promoter are clearly distinct from those of the rpL30 and rpL32 promoters.

Contact residues for δ -factor binding

The relative novelty of internal sequence elements in polymerase II promoters prompted us to investigate the

Figure 5. Specificity of rpL30 factor binding and competition with other ribosomal protein gene fragments. (Top) Results of a survey of binding sites in the rpL30 promoter region by the gel mobility-shift assay. Strong binding sites were localized in the segments designated α , β , γ , and δ . No binding was observed in the segments marked 0. (A) rpL30- α ; (B) rpL30- β ; (C) rpL30- γ ; (D) rpL30- δ . The labeled rpL30 fragments were incubated without and with nuclear extract from S194 mouse plasmacytoma cells (lanes 1 and 2, respectively) or with extract and a 100 $\times$ molar excess of various unlabeled competitor DNA fragments (lanes 3, 4, etc.), electrophoresed on native 5% polyacrylamide gels, and visualized by autoradiography. (Bottom) The nature of the unlabeled competitor and the outcome of the competition experiments.

nature of δ -factor binding in greater detail. Methylation interference analysis (Siebenlist and Gilbert 1980) was used to identify the guanine residues that constitute δ -factor contact points on both the sense and antisense strands of rpL30 (Fig. 6). Six contact residues were identified in the +15 to +22 region. A similar analysis of the rpL32- δ -factor site identified seven contact residues in the +28 to +37 region (Moura-Neto et al. 1989 and data not shown). Both δ sites are entirely within the first (noncoding) exons of these genes and both share the common sequence 5'-CNGCCATCT-3'.

The importance of δ -factor binding for rpL30 promoter activity

To determine whether δ -factor binding is critical for rpL30 promoter activity, we used a site-specific mutagenesis protocol to generate two variants of the wild-type L30- δ -binding site in construct g (Fig. 7A). One variant, g1, has substitutions at six positions and retains only one of the six original contact residues. The other variant, g2, has substitutions at two positions and retains four contact residues. DNA fragments containing the g1 variant site lack the capacity to bind the δ factor, whereas those containing the g2 variant still retain this capacity (Fig. 7B). Correspondingly, rpL30 promoters bearing the g1 site are <10% as active as the wild-type rpL30 promoter g, whereas those bearing the g2 site are

almost equally active (Fig. 7C). These results clearly demonstrate that the δ -factor binding is critical for proper functioning of the rpL30 promoter and further emphasize the importance of internal sequence elements for the transcriptional activity of ribosomal protein genes.

Isolated ribosomal protein promoters are equipotent in driving CAT expression

Having established the boundaries of the elements required for maximum activity of the rpL30, rpL32, and rpS16 promoters, we were in a position to examine their relative strength in driving the expression of a common reporter gene. Accordingly, we compared the expression in mouse plasmacytoma cells of the rpL30-CAT construct g, an rpS16-CAT construct containing the -179 to +29 region of rpS16 (c-CAT, Hariharan and Perry 1989), and an rpL32-CAT construct containing the -159 to +115 region of rpL32 linked to a functional 3'-splice junction. The CAT activity of these three constructs was remarkably uniform in two independent transfection experiments ($\pm 13\%$ variance when standardized to the amount of plasmid DNA in Hirt supernatant fractions of transfected cells). To the extent that the relative CAT activity reflects relative transcriptional efficiency, this result supports our supposition that the equivalency of ribosomal protein gene transcription

Figure 6. Methylation interference analysis of δ -factor binding to the rpl30 promoter. Appropriate rpl30 fragments, ^{32}P -labeled in the sense or antisense strand, were partially methylated with dimethylsulfate and incubated with S194 nuclear extract. The bound (B) and unbound (UB) fragments were separated on polyacrylamide gels, cleaved with piperidine, and analyzed on DNA sequencing gels, together with (CA)-cleaved and G-cleaved markers. The positions of the uncleaved G residues in the bound DNA are indicated by arrows. (Bottom) Individual G residue contact points are circled in the sequence.

rates, demonstrated by the nuclear run-on analysis, stems from the equipotency of their promoters.

Discussion

Common features of mouse ribosomal protein promoters

Our analysis of the mouse rpl30 promoter has revealed several novel properties that are shared to different extents by other ribosomal protein promoters. The three mouse ribosomal protein promoters that have been examined in detail, rpl30, rpl32, and rpl16, all consist of multiple (five or more) elements distributed over a stretch of ~ 200 bp (Fig. 4). A remarkable feature of these ribosomal protein promoter elements is that the array invariably extends downstream of the cap site to include portions of the first exon and, in two cases, rpl30 and rpl32, adjacent regions of the first intron. This feature,

together with an unusual cap site that is embedded in a polypyrimidine tract (Dudov and Perry 1984, 1986), sets the ribosomal protein promoters apart from all other polymerase II promoters that have been examined to date. Another common feature of these ribosomal protein promoters is that the upstream elements contribute to, but are not absolutely essential for, promoter activity. The rpl30 and rpl32 promoters operate at $\sim 25\%$ of maximum efficiency with the internal elements and only 36 bp of 5'-flanking sequence. The rpl16 promoter has $\sim 40\text{--}45\%$ of maximum efficiency when only 83 bp of 5'-flanking sequence are present. Interestingly, the promoter cores that suffice for basal activity, elements C, D, and E for rpl30 and rpl32 and elements B/B', C, and D for rpl16, cover a stretch of ~ 110 bp in all three ribosomal protein genes.

In drawing the foregoing conclusions, we have presumed that the relative activities of the various deletion mutants are primarily attributable to the retention or loss of particular elements and not to alterations in the sequence context of the residual elements. In general, this premise seems justified because of the excellent correlation between the location of functionally important elements and factor-binding sites. For the few cases in which a correlation is not evident, a contextual effect could conceivably be a contributing factor.

The resemblance between the promoter structures of rpl30 and rpl32 is particularly striking. In these promoters, each of the similarly located modules: B, C, and D, interacts with a nuclear factor (β , γ , δ , respectively), which has an indistinguishable binding specificity for the corresponding elements of both genes. Because these three modules have the same spatial relationship in the two genes, it seems reasonable to suppose that a common set of factors may be used by both promoters. The factors used by the rpl16 promoter are clearly distinct from the rpl30/rpl32 factors, as indicated by the lack of any cross-competition for factor binding. A comparison of the sequences that comprise this region of the ribosomal protein promoters is presented in Figure 8.

The rpl30/rpl32- δ factor-binding sites were identified by methylation interference analysis. Both sites embrace the sequence 5'-C_GC_GC_GC_GCCATCT-3'. Interestingly, the sequence 5'-GGCGGCCATC-3' is also present in an adjacent intron segment of rpl32, which effectively competes with both rpl30 and rpl32 exon I segments for δ -factor binding and which, itself, binds a factor that was previously termed ϵ (Atchison et al. 1989). We suspect that δ and ϵ may be the same factor and, therefore, that rpl32 actually contains two δ -factor sites separated by ~ 32 bp or three helical turns. There is no comparable sequence in the corresponding intron 1 region of rpl30, and fragments containing this region neither compete for δ -factor binding nor bind to a nuclear factor under the conditions of our mobility-shift assay. This latter observation was unexpected because functional analyses indicated that this intron 1 region (E) is important for the activity of both rpl30 and rpl32 promoters. If an rpl30 E region-specific factor exists, it is clearly different from δ and is either routinely absent from our nuclear extracts

residues within the site is sufficient to abolish binding. Presumably, the interaction of δ factor with its binding sites is more sensitive to the presence of bulky methyl groups in the DNA major groove than to the replacement of two GC base pairs by TA base pairs. When the chemical nature of the DNA-binding domain of δ factor is better understood, this observation should help in the formulation of a reasonably detailed model of the DNA-factor interaction.

The rpl32- β -factor site was previously localized by differential gel-mobility shift assays to an 11-bp segment between positions -79 and -69, which contains the sequence 5'-CCGGAAGTG-3' (Atchison et al. 1989). A very closely related sequence, 5'-CCGGAAGCG-3', occurs in an inverted orientation in the segment that contains the rpl30- β -factor site. This common sequence could account for the effective cross-competition between rpl30 and rpl32 fragments for β -factor binding.

The segments of rpl30 and rpl32 that contain the γ -factor binding site both contain the sequence RCYTTGCR at positions -23 to -16 and -26 to -19, respectively, and a recent methylation interference analysis (N. Hariharan, unpubl.) has confirmed that both γ -factor binding sites embrace this sequence. Although the similarity between the rpl30 and rpl32 γ sites is not as striking as that of the δ - and β -factor sites, the recognition of an identical transcription factor by apparently dissimilar sequences is not unprecedented (Pfeifer et al. 1987; Hwung et al. 1988).

The majority of mammalian genes transcribed by polymerase II have a TATA motif in the region corresponding to the C elements of the ribosomal protein genes. In the murine ribosomal protein genes this motif is particularly degenerate, the -30 to -24 sequence being TAGAAGA in rpl30, TCATACC in rpl32, and GAAAAAT in rps16. Despite the degeneracy of these motifs, the transcriptional initiation sites are selected with very high precision, as evidenced by the uniqueness of the ribosomal protein cap sites (Dudov and Perry 1984; Wiedemann and Perry 1984; Wagner and Perry 1985). Thus, the ribosomal protein promoters as a class do not exhibit the imprecise start-site selection that is frequently associated with *pol* II genes that lack canonical TATA motifs (see Reynolds et al. 1984; Saffer and Singer 1984; Singer-Sam et al. 1984, and references quoted therein). The role of the γ factor may be analogous to that of the TATA-box-specific factor TFIID (Nakajima et al. 1988), which is believed to participate in the correct positioning of polymerase II.

Importance of internal elements for ribosomal protein promoter activity

Most regulatory elements that are located within the transcribed portion of a gene have the characteristics of an enhancer, i.e., they function over large distances and are both orientation and position independent. In contrast, the internal elements of the ribosomal protein genes seem to be integral parts of the promoter structure. They are located within 80 bp of the start site and are strongly position dependent (Chung and Perry 1989).

Several lines of evidence indicate that these internal ribosomal protein elements act at the transcriptional level. First, the exon I elements (D) of all three ribosomal protein genes have been shown to bind specific nuclear factors. In this study we have demonstrated that the binding of the rpl30/rpl32- δ factor to its recognition site in rpl30 exon I is essential for proper rpl30 promoter function. Binding site-specific base substitutions

that abolish the δ -factor interaction greatly decrease promoter activity, whereas binding-site substitutions that do not affect δ -factor interaction are innocuous. Second, the importance of the rpl32 intronic element (E) for transcriptional efficiency was directly demonstrated by nuclear run-on analysis (Chung and Perry 1989). As argued above, this element probably also interacts with the rpl30/rpl32- δ factor. Third, the effect of deleting the internal elements was basically the same whether the promoters were in their natural positions in the ribosomal protein genes or linked to the CAT reporter gene [Atchison et al. 1989; Chung and Perry 1989; Moura-Neto et al. 1989; this paper (constructs f versus m)]. If the deletions had affected mRNA stabilities rather than transcription rates, one would not anticipate such similar results with the disparate ribosomal protein and CAT transcripts.

The relationship of these internal promoter elements to the polypyrimidine-embedded cap sites of the ribosomal protein genes is still unknown. The polypyrimidine stretch may be required to confer a unique conformational property on the DNA that separates the γ - and δ -factor binding sites. The widespread occurrence of this unusual feature among the vertebrate ribosomal protein genes (see Loreni et al. 1985; Beccari and Mazzetti 1987; Chen and Roufa 1988, and references quoted therein) leads us to suspect that it has an important role in ribosomal protein gene expression.

To the best of our knowledge, only a few other mammalian promoters are known to have positive transcriptional elements located immediately downstream of the cap site (for references, see Jones et al. 1988), and none of these bear any striking resemblance to the ribosomal protein promoters. The mouse *c-myc* gene has a positive element at the 3' end of exon I (Yang et al. 1986), a region that also contains a binding site for the rpl30/rpl32- δ factor (Atchison et al. 1989). However, this region is ~500 bp downstream of the cap site. The internal elements of the human immunodeficiency virus promoters, which are very near to the cap site, contain binding sites for at least two cellular transcription factors. The recognition sequences of these factors appear to be unrelated to those of the ribosomal protein elements.

Equipotency of ribosomal protein promoters

Although the various ribosomal protein promoters are not composed of exactly the same set of modules, the overall organization of the modules is similar, so that each combination creates a promoter that appears to have a remarkably similar efficiency in a wide variety of cell types. The similarity in promoter strength is indicated by the equivalent polymerase II loading of the different ribosomal protein genes, as judged from nuclear run-on experiments, and the equal efficiency of the isolated promoters in driving the expression of a CAT reporter gene. This uniformity in transcription rate is ultimately responsible for the relatively uniform abundance of the ribosomal protein mRNAs and the balanced syn-

thesis of ribosomal proteins. In mammalian cells, the short-term regulation of ribosomal protein production in response to environmental or physiological changes seems to be at the level of ribosomal protein mRNA translation or of the assembly and turnover of ribosomal proteins (for references, see Meyuhas et al. 1987). The long-term coordination of ribosomal protein synthesis is apparently achieved by the uniform intrinsic strength of the ribosomal protein promoters and the relatively uniform stability and translational efficiency of the ribosomal protein mRNAs.

Evolution of ribosomal protein promoters

A striking similarity in the promoter structure of genes encoding different ribosomal proteins has also been observed in yeast (for references, see Vignais et al. 1987). In this case, the critical elements, a set of generally duplicated motifs, termed UAS_{rp} , are located several hundred base pairs upstream of the transcriptional start site. Despite some sequence variation among these motifs, most of them appear to bind the same transcription factor, TUF. Thus, the promoters of mammalian and yeast genes encoding a particular ribosomal protein are characteristic of the species of organism and totally different from each other. This is dramatically illustrated by a comparison of the yeast and mouse genes encoding rpL30 (Fig. 9). The central protein-encoding region is highly conserved, exhibiting only 37.5% nucleotide sequence divergence in over 1.1 billion years of evolution (Dabeva and Warner 1987; Field et al. 1988). This extremely high degree of conservation is characteristic of genes that encode the protein constituents of complex particles in which there are very large constraints for proper assembly. In marked contrast, the sequences flanking this region, including the promoter segments, are completely dissimilar. Nevertheless, each of these promoters has features in common with other ribosomal protein promoters of the species.

The fact that several relatively novel features are shared by the promoters of evolutionarily unrelated ribosomal protein genes seems to suggest a form of convergent evolution in which different assortments of

basic elements are organized and selected for their combined performance in eliciting comparable promoter efficiencies. This selection is presumably subject to various constraints, such as the kinds of transcription factors that are available to a particular organism or the necessity to function in diverse cell types, that are all determined by a single genome. Evolutionary pathways involving both the piecemeal addition of individual elements and the duplication of element assemblies may be envisioned. This could account for the greater similarity of some ribosomal protein promoters, e.g., rpL30 and rpL32, than of others, e.g., rpL30 and rpS16.

Materials and methods

The procedures for cell growth, nuclear run-on analysis, DNA transfection, RNA analysis by S1 nuclease protection, CAT assays, and gel-mobility shift assays have been described in previous publications from this laboratory (Kelley and Perry 1986; Chung and Perry 1989; Hariharan and Perry 1989). For the run-on experiments, we used denatured DNA segments spanning the ribosomal protein genes (Fig. 1) and the J-C region of the κ immunoglobulin gene [PEC κ (Coleclough et al. 1981)]. A variety of evidence, including S1 nuclease protection analyses of nuclear RNA with + and - strand probes and Northern blot analyses of nuclear RNA with denatured probes, indicates that there is only + strand transcription from the ribosomal protein and κ genes. For the S1 nuclease protection analyses, we used a 266-bp *Pst*I-*Mst*II rpL30 probe derived from construct f (see below) and a 167-bp rpS16 probe, which yields a 68-nucleotide fragment when protected by properly initiated rpS16 transcripts (Hariharan and Perry 1989).

For the methylation interference analyses, we used the following 5'-end labeled fragments: rpL30, -31 to +49 for the sense strand and -52 to +49 for the antisense strand; rpL32, -36 to +72 for both strands. These fragments were partially methylated with dimethylsulfate (Siebenlist and Gilbert 1980), incubated with nuclear extract, and the bound and unbound fragments were separated by electrophoresis on 5% polyacrylamide gels. The most intense retarded band and the unretarded band were eluted, extracted with phenol/chloroform, and precipitated with ethanol. This DNA was treated with 0.5 M piperidine and electrophoresed on an 8% polyacrylamide gel containing 50% urea.

The various rpL30 constructs were made by standard recombinant DNA procedures. A *Pst*I-*Bgl*II fragment from the ge-

Figure 9. A comparison of the rpL30 genes in *Saccharomyces cerevisiae* and *Mus musculus*. The exons are shown as large boxes, with the translation initiation and termination codons marked by open and solid inverted triangles, respectively. Transcriptional start sites are shown as horizontal arrows. The promoter elements are schematically diagrammed as solid boxes (R). The *S. cerevisiae* data are from Dabeva and Warner (1987); the *M. musculus* data are from Wiedemann and Perry (1984) and this paper.

nomic clone rpl30-1 (Wiedemann and Perry 1984), which extends from position -205 to 260 bp beyond the poly(A) addition site, was inserted into the *Pst*I-*Bam*HI sites of pUC18 to obtain construct a. Constructs b-e were derived from construct a by a combination of cleavage at various restriction sites, trimming or end-filling with the Klenow enzyme, and blunt or cohesive-end ligations. The sites used for the 5' termini of these constructs were: *Hha*I (-94), *Ava*II (-52), *Mn*II (-31), and *Sac*I (-9).

Construct f, which lacks intron 1, was made by substituting a *Bgl*II-*Mst*II fragment of construct a, which extends from +21 to +167, with a corresponding fragment from a highly conserved processed pseudogene L30-2 (Wiedemann and Perry 1984). In constructs g, i, k, and m, which were derived from a, b, c, and f, respectively, the *Mst*II site at +167 was end-filled and ligated to the blunt-ended *Xba*I site of a plasmid bearing a promoterless CAT gene, p106 [Atchison et al. 1989]. Constructs h and j were similarly produced with fragments that had 5' termini at -135 and -75 (*Ava*II sites). Constructs l, n, o, and p were derived from construct g by appropriate cleavage, trimming or filling, and ligation. The restriction sites used for the 3' termini of these constructs were *Bgl*II (+20), *Hph*I (+49), *Ava*I (+77), and *Nhe*I (+92).

In constructs n-p, a 73-bp *Nco*I-*Nla*IV fragment from rpS16 was inserted between the rpl30 promoter segment and the CAT gene. This fragment, which contains 65 bp of intron 2 and 8 bp of exon III, provides a functional 3'-splice site that can interact with the intron-1 5'-splice junction of rpl30. In preliminary experiments we observed that a 3'-splice site in this location was essential to obtain maximum expression from rpl30-CAT constructs that contained the intron-1 5'-splice junction. The CAT activity of constructs containing this 3' site was about two- to threefold greater than that of constructs lacking the site. A similar effect was observed when a 3'-splice site was supplied to an unpaired intron-1 5'-splice junction in an rpl32-CAT construct. Downstream from the CAT gene in p106, there is an SV40 segment that contains a spliceable intron and a poly(A) addition site. We assume that in the absence of the added 3' sites, splicing may occur between the unpaired 5' junctions and the 3' splice site of the SV40 intron. Such a splice would eliminate the CAT-encoding sequence and thus reduce the amount of CAT activity per unit transcript.

For constructs g1 and g2, the -4 (*Sac*I) to +20 (*Bgl*II) segment of construct g was replaced by synthetic oligonucleotides that contained the substitutions shown in Figure 7, as well as the appropriate restriction site joints. The rpl32-CAT gene was made by substituting the -205 to +90 region of construct g with a -159 to +115 fragment that was isolated from a previously described rpl32 construct [Atchison et al. 1989]. The rpS16-CAT gene is described elsewhere [Hariharan and Perry 1989].

Acknowledgments

We are especially indebted to Michael Atchison and Shan Chung for stimulating discussions and exchange of both theoretical and technical information on ribosomal protein promoter analysis. This research was supported by grants from the National Science Foundation (DCB84-13609), the National Institutes of Health (AI-17330, CA-06927, RR-05539), and an appropriation from the Commonwealth of Pennsylvania.

References

Atchison, M.L., O. Meyuhas, and R.P. Perry. 1989. Localization of transcriptional regulatory elements and nuclear factor

- binding sites in the mouse ribosomal protein gene rpl32. *Mol. Cell. Biol.* **9**: 2067-2074.
- Beccari, E. and P. Mazzetti. 1987. The nucleotide sequence of the ribosomal protein L14 gene of *Xenopus laevis*. *Nucleic Acids Res.* **15**: 1870.
- Chen, I.-T. and D.J. Roufa. 1988. The transcriptionally active human ribosomal protein S17 gene. *Gene* **70**: 107-116.
- Chung, S. and R.P. Perry. 1989. Importance of introns for expression of the mouse ribosomal protein gene rpl32. *Mol. Cell. Biol.* **9**: 2075-2082.
- Coleclough, C., R.P. Perry, K. Karjalainen, and M. Weigert. 1981. Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature* **290**: 372-378.
- Dabeva, M.D. and J.R. Warner. 1987. The yeast ribosomal protein L32 and its gene. *J. Biol. Chem.* **262**: 16055-16059.
- Dudov, K. and R.P. Perry. 1984. The gene family encoding the mouse ribosomal protein L32: Characteristics of the uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* **37**: 457-468.
- . 1986. Properties of a mouse ribosomal protein promoter. *Proc. Natl. Acad. Sci.* **83**: 8545-8549.
- Field, K.G., G.J. Olsen, D.J. Lane, S.J. Giovannoni, M.T. Ghiselin, E.C. Raff, N.R. Pace, and R.A. Raff. 1988. Molecular phylogeny of the animal kingdom. *Science* **239**: 748-752.
- Fried, M.G. and D.M. Crothers. 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**: 6505-6525.
- Hariharan, N. and R.P. Perry. 1989. A characterization of the elements comprising the promoter of the mouse ribosomal protein gene rpS16. *Nucleic Acids Res.* **17**: 5323-5337.
- Hwung, Y.-P., L.-H. Wang, S.Y. Tsai, and M.-J. Tasi. 1988. Differential binding of the chicken ovalbumin upstream promoter (COUP) transcription factor to two different promoters. *J. Biol. Chem.* **263**: 13470-13474.
- Jones, K.A., P.A. Luciw, and N. Duchange. 1988. Structural arrangements of transcription control domains within the 5'-untranslated leader regions of the HIV-1 and HIV-2 promoters. *Genes Dev.* **2**: 1101-1114.
- Kelley, D.E. and R.P. Perry. 1986. Transcriptional and posttranscriptional control of immunoglobulin mRNA production during B lymphocyte development. *Nucleic Acids Res.* **14**: 5431-5447.
- Lindell, T.J., F. Weinberg, P.W. Morris, R.G. Roeder, and W.J. Rutter. 1970. Specific inhibition of nuclear RNA polymerase II by α -amanitin. *Science* **170**: 447-449.
- Lorenzi, F., I. Ruberti, I. Bozzoni, P. Pierandrei-Amaldi, and F. Amaldi. 1985. Nucleotide sequence of the L1 ribosomal protein gene of *Xenopus laevis*: Remarkable sequence homology among introns. *EMBO J.* **4**: 3483-3488.
- Meyuhas, O. and R.P. Perry. 1980. Construction and identification of cDNA clones for several mouse ribosomal proteins: Application for the study of r-protein gene expression. *Gene* **10**: 113-129.
- Meyuhas, O., E.A. Thompson, and R.P. Perry. 1987. Glucocorticoids selectively inhibit translation of ribosomal protein mRNAs in P1798 lymphosarcoma cells. *Mol. Cell. Biol.* **7**: 2691-2699.
- Moura-Neto, R., K.P. Dudov, and R.P. Perry. 1989. An element downstream of the cap site is required for transcription of the gene encoding mouse ribosomal protein L32 gene. *Proc. Natl. Acad. Sci.* **86**: 3997-4001.
- Nakajima, N., M. Horikoshi, and R.G. Roeder. 1988. Factors involved in specific transcription by mammalian RNA polymerase II: Purification, genetic specificity, and TATA box-

- promoter interactions of TFIID. *Mol. Cell. Biol.* **8**: 4028–4040.
- Pfeifer, K., T. Prezant, and L. Guarente. 1987. Yeast HAP1 activator binds to two upstream activation sites of different sequence. *Cell* **49**: 19–27.
- Reynolds, G.A., S.K. Basu, T.F. Osborne, D.J. Chin, G. Gil, M.S. Brown, J.L. Goldstein, and K.L. Luskey. 1984. HMG CoA reductase: A negatively regulated gene with unusual promoter and 5' untranslated regions. *Cell* **38**: 275–285.
- Saffer, J.D. and M.F. Singer. 1984. Transcription from SV40-like monkey DNA sequence. *Nucleic Acids Res.* **12**: 4769–4788.
- Siebenlist, U. and W. Gilbert. 1980. Contacts between *Escherichia coli* RNA polymerase and an early promoter of the phage T7. *Proc. Natl. Acad. Sci.* **77**: 122–126.
- Singer-Sam, J., D.H. Keith, K. Tani, R.L. Simmer, L. Shively, S. Lindsay, A. Yoshida, and A.D. Riggs. 1984. Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate kinase. *Gene* **32**: 409–417.
- Vignais, M.-L., L.P. Woudt, G.M. Wassenaar, W.H. Mager, A. Sentenac, and R.J. Planta. 1987. Specific binding of TUF factor to upstream activation sites of yeast ribosomal protein genes. *EMBO J.* **6**: 1451–1457.
- Wagner, M. and R.P. Perry. 1985. Characterization of the multigene family encoding the mouse S16 ribosomal protein: Strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. *Mol. Cell. Biol.* **5**: 3560–3576.
- Wiedemann, L.M. and R.P. Perry. 1984. Characterization of the expressed gene and several processed pseudogenes for the mouse ribosomal protein L30 gene family. *Mol. Cell. Biol.* **4**: 2518–2528.
- Wiedemann, L.M., P. D'Eustachio, D.E. Kelley, and R.P. Perry. 1987. Three functional ribosomal protein genes are unlinked in the mouse genome. *Somatic Cell Mol. Genet.* **13**: 77–80.
- Yang, J.-Q., E.F. Remmers, and K.B. Marcu. 1986. The first exon of the *c-myc* proto-oncogene contains a novel positive control element. *EMBO J.* **5**: 3553–3562.

Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements.

N Hariharan, D E Kelley and R P Perry

Genes Dev. 1989, **3**:

Access the most recent version at doi:[10.1101/gad.3.11.1789](https://doi.org/10.1101/gad.3.11.1789)

References

This article cites 30 articles, 13 of which can be accessed free at:
<http://genesdev.cshlp.org/content/3/11/1789.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).