Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1526–1530, February 1990
Biochemistry

Functional dissection of a mouse ribosomal protein promoter: Significance of the polypyrimidine initiator and an element in the TATA-box region

(transcription efficiency/initiation complex)

NARAYANAN HARIHARAN AND ROBERT P. PERRY

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111

Contributed by Robert P. Perry, December 15, 1989

ABSTRACT **All of the mammalian ribosomal protein (rp) genes examined to date initiate transcription with high precision despite the fact that they do not contain a well-defined TATA box. The initiation sites are situated within polypyrimidine tracts that are flanked by both upstream and intragenic promoter elements. In the TATA-box region of each rp promoter, there is a functionally critical element with nuclear factor binding specificity that is distinct from that of a conventional TATA box. To understand how the various elements contribute to rp promoter function, we have used site-specific mutagenesis–transfection protocols and factor binding analyses to evaluate the significance of the polypyrimidine initiator and the TATA-box counterpart for efficient and accurate transcription of the rpS16 gene. Our results indicate (*i*) that the polypyrimidine initiator sequence critically defines the position of the transcriptional start site, whereas a much less specific sequence is sufficient to satisfy the efficiency requirement; (*ii*) that an uninterrupted stretch of pyrimidines in the initiator region is not necessary for efficient transcription of rpS16 gene; and (*iii*) that the TATA-box counterpart or even a substituted conventional TATA box primarily influences promoter efficiency. The great diversity of promoter design, which is becoming evident as more RNA polymerase II promoters are being carefully dissected, suggests that the requirements for building a functional initiation complex may be much more flexible than was previously appreciated.**

Promoters perform two major functions for genes transcribed by RNA polymerase II. They define the site of transcriptional initiation (cap site) and help determine the efficiency of polymerase loading at that site. These functions are attributed to an array of cis-acting elements, which serve as binding sites for protein factors that help the polymerase to form a transcriptional initiation complex (1). In higher eukaryotes, the promoter core usually contains a TATA box or some functional counterpart, located 20–30 base pairs (bp) upstream of the cap site, and an initiator element, which encompasses the cap site. The relative importance of these two elements for proper promoter function seems to vary widely among different promoters. Among genes that initiate transcription at one or a few closely clustered sites, there are examples such as the herpesvirus thymidine kinase promoter and the β-globin promoter, in which TATA-box interactions are apparently more critical than initiator interactions (2, 3), examples such as the adenovirus major late promoter and the fibroin promoter, which appear to require both TATA-box and initiator interactions (4, 5), and examples such as the simian virus 40 late promoter and the terminal deoxynucleotidyltransferase promoters, which lack discernible TATA elements and depend heavily on initiator interactions (6, 7). Many other TATA-box-lacking genes exhibit imprecise initiation (8); conceivably, such genes lack effective initiator elements.

The promoters of mammalian ribosomal protein (rp) genes initiate transcription with high precision despite the fact that they do not contain well-defined TATA boxes (ref. 9 and references therein). The initiation sites are generally situated within polypyrimidine tracts flanked by regions of high G+C content. A detailed comparison of three mouse rp promoters has revealed a common architecture in which essential elements are located both upstream and downstream of the cap site (9). In the TATA-box region of each promoter, there is a functionally critical element which contains a nuclear factor binding site. It is of considerable interest to know how these various elements contribute to rp promoter function. To this end, we have used site-specific mutagenesis–transfection protocols to evaluate the significance of the polypyrimidine initiator and the TATA-box counterpart for efficient and accurate transcription of the rpS16 gene. Our results indicate that the initiator has an important role in defining the position of the transcriptional start site and that the element in the TATA-box region primarily influences promoter efficiency. It is also evident that an uninterrupted stretch of pyrimidines in the initiator region is not necessary for efficient transcription of the rpS16 gene.

MATERIALS AND METHODS

Various rpS16 mutants were made by standard cloning procedures from a construct (*c*, ref. 10) that contains 179 bp of 5′ flanking sequence, the entire transcribed portion of the gene, and 107 bp of 3′ flanking sequence. This construct contains all of the elements needed for efficient expression of the rpS16 gene (10) (Fig. 1*A*). Synthetic oligonucleotides representing the −37 to −12 region or the −11 to +29 region and containing point mutations in the −30 to −24 or −4 to +8 region were substituted for the corresponding wild-type sequences in the above construct by directional cloning protocols. The mutations were confirmed by sequencing through the altered regions. The procedures for cell growth, DNA transfection, RNA analysis by S1 nuclease protection, gel-mobility-shift assays, and methylation interference analysis have been described elsewhere (9–11). For the methylation interference analysis, the modification and cleavage reactions were done as described by Siebenlist and Gilbert (12) so that the interference to binding by A methylation could be detected. For the S1 nuclease protection assays of rpS16 expression, 167-bp DNA probes spanning the cap sites of wild-type and mutant genes were used (Fig. 1*A*). The wild-type probe yields a 68-nucleotide band when protected by a correctly initiated transcript (10). Transcripts from cotransfected rpL32 genes were detected with a 243-bp DNA

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

Abbreviation: rp, ribosomal protein.

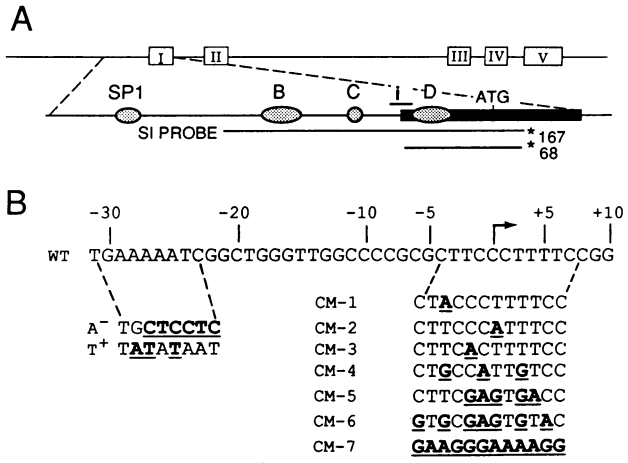

FIG. 1. Structure of the wild-type and mutant rpS16 promoters. (*A*) Beneath the diagram of the complete gene (exons labeled with roman numerals) is an expansion of the promoter region showing the locations of four elements (stippled ovals), the initiator segment (i), and the first exon (solid bar) with the position of the ATG initiator codon indicated. The S1 nuclease probe and protected fragment are schematically diagrammed below. (*B*) Sequences of the −31 to +10 region and the various substitution mutants.

probe that yields an 86-nucleotide fragment when protected by a correctly initiated and properly spliced transcript (11).

RESULTS

Effect of Purine Substitutions in the Polypyrimidine Initiator. To evaluate the significance of the novel polypyrimidine initiator element, we constructed a series of mutant rpS16 genes in which one or more pyrimidines in the element were replaced by purines (Fig. 1*B*). This series included single substitutions upstream of, downstream of, and directly at the cap site (CM.1, CM.2, and CM.3, respectively), a triple substitution that reduced the number of consecutive pyrimidines to three or less (CM.4), more drastic mutations containing five and eight substitutions (CM.5 and CM.6, respectively) and an antistrand substitution (CM.7) in which the entire polypyrimidine tract was replaced by its complementary polypurine sequence. These mutant constructs were transfected into COS.7 cells and their transient expression was assayed by S1 nuclease protection analysis of cytoplasmic RNA. An intact rpL32 gene was cotransfected to control for transfection efficiency and RNA yield. Each RNA sample was assayed both with a common S1 probe derived from the −99 to +67 region of CM.6 and with individual S1 probes derived from the corresponding regions of each of the variant rpS16 genes. Assays with a common S1 probe are advantageous for quantitatively measuring the relative level of expression of the mutant constructs, whereas assays with individual probes enable us to evaluate effects on the positioning of the transcriptional start-site.

Assays with the common probe (Fig. 2*A*) demonstrated that an interruption in the continuity of the polypyrimidine tract by single or even three purine substitutions has no detectable effect on the level of rpS16 expression. The amount of protected fragment observed with the CM.1, CM.2, CM.3, and CM.4 mutants was essentially identical to that obtained with the wild-type gene. In contrast, the more extensive purine substitutions caused a marked reduction in expression, as shown by the results with the CM.5, CM.6, and CM.7 mutants. The amount of protected fragment observed with these mutants was about 20–25% of the wild-type level. Since the transcriptional start sites are all upstream of +8 (see below), the sizes of the various protected fragments in this assay mainly reflect the extent of homology between the CM.6 probe and the wild-type or mutant transcripts. Accordingly, the fragment sizes are about 60 nucleotides for the wild-type, CM.1, CM.2, CM.3, CM.4, and CM.7 constructs (+8/9 to +67 homology), 68 nucleotides for CM.6 (full homology), and 60 and 68 nucleotides for CM.5 (partial homology). These data, summarized in Table 1, lead us to conclude that a polypyrimidine tract is not required for efficient function of the rpS16 promoter. Nevertheless, the initiator element is clearly an important contributor to promoter function, as evidenced by the fact that it cannot tolerate gross sequence changes without loss of activity.

Assays with the individual S1 probes (Fig. 2*B*) revealed some differences in the position of the start site among the various constructs. For the most part, these sites were within a few base pairs of the normal cap site. However, in the case of the more drastic mutants, CM.5 and CM.6, we could also detect an array of less abundant larger fragments, indicating that some transcripts were being initiated 10 to 30 bp further upstream. To determine the major start sites more accurately, we analyzed the S1 nuclease-protected products on a high-resolution polyacrylamide gel together with a DNA sequence ladder that spans the cap site (Fig. 3). Previous calibration of such gels (13, 14) has shown that fragments protected by full-length mRNA migrate 2.5 to 4.5 nucleotides slower than the cap site nucleotide in a Maxam–Gilbert DNA sequence ladder, the heterogeneity being most likely due to

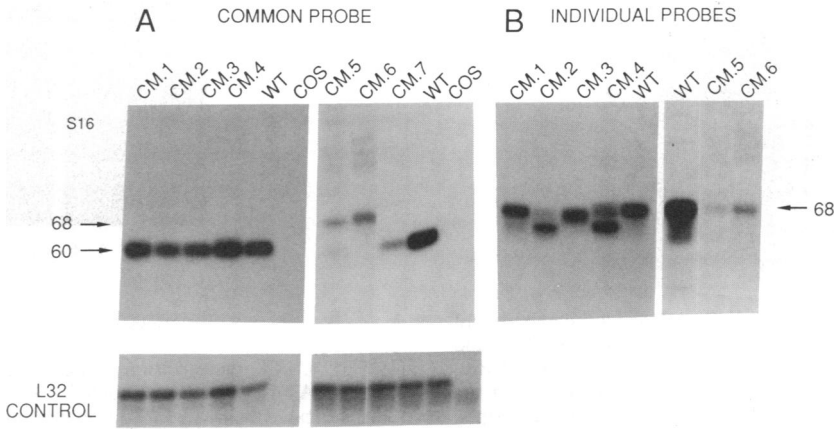

FIG. 2. Expression of wild-type (WT) and initiator mutant (CM series) rpS16 genes. S1 nuclease protection assays were carried out with the CM.6 probe (*A*) and with individual probes derived from the wild type and each of the mutant genes (*B*). The arrows indicate the 60- and 68-nucleotide protected fragments. Lanes marked COS are assays of RNA from untransfected cells. The lower left gels are parallel S1 assays of a cotransfected wild-type rpL32 gene.

Table 1. Efficiency and accuracy of rpS16 mutant constructs

Mutant	Relative efficiency*	Start sites†	
		Major	Minor
Wild type	100	+1 (C)	
CM.1	100	+3 (T)	+2 (C)
CM.2	100	+7 (C)	+2 (C)
CM.3	100	+3 (T)	+2 (C)
CM.4	100	+6 (T)	−1, +2 (C, A)
CM.5	20	+3 (G)	
CM.6	20	+3 (G)	+1 (G)
CM.7	20		ND
A^-	10	+1 (C)	
T^+	100	+1 (C)	
T^+CM.2	100	+7 (C)	+2 (C)
T^+CM.3	100	+3 (T)	+2 (C)
T^+CM.4	100	+6 (T)	−1, +2 (C, A)
T^+CM.6	20	+3 (G)	+1 (G)

*Efficiencies determined from densitometric scans of the experiments of Figs. 2*A*, 5*A*, and 5*D* and replicate experiments.

†Start sites determined from the data of Figs. 3 and 5*B* and similar high-resolution analyses of the experiments shown in Fig. 6 *A* and *B*. The major start site was taken to be the sequence position that is 4.5 nucleotides below the largest of the most abundant S1-resistant fragments. Minor sites were considered when the fragment corresponding to the major site was not the largest detectable fragment. In this case, the largest fragments of intermediate or lower abundance were used to position the minor sites (see dots in Fig. 3). ND, not determined.

steric hindrance of S1 nuclease by the cap structure. This relationship is clearly illustrated in Fig. 3 by a comparison of the wild-type S1 products and a sequence ladder of the CM.3 initiator region.

When the wild-type and mutant S1 patterns were compared with regard to both size and relative abundance (intensity) of the fragments, it was evident that the position of the start site was altered in all of the mutants. It is noteworthy that even the subtle mutations CM.1 through CM.4, which have no effect on promoter efficiency, shift the start site downstream from its normal position and decrease the precision of the initiation event. Interestingly, all except one of the start sites selected by these subtle mutants correspond to pyrimidine residues, whereas for mutants CM.5 and CM.6, the residual transcription initiates at a purine (Table 1). These results demonstrate that even minor alterations in the initiator sequence can affect the positioning of the initiation complex, thus confirming the importance of this element for rpS16 promoter function.

Characterization of the Element in the TATA-Box Region. Previously we demonstrated that the −37 to −12 region of the rpS16 gene contains both an element that is essential for efficient expression and a specific binding site for a nuclear factor (10). A subsequent methylation interference analysis indicated some weak interference within the A_5 segment at −29 to −25, suggesting that this segment may be involved in the binding reaction. Since this location is frequently occupied by a TATA box in other RNA polymerase II genes, we suspected that the A_5 sequence might either be a divergent motif that has retained the ability to interact with the conventional TATA-box factor, TFIID (15), or, alternatively, be part of a binding site for a factor that is a functional equivalent of TFIID.

To investigate these possibilities we constructed two mutants, A^- and T^+, in which the wild type −31 to −24 sequence TGAAAAAT was specifically altered (Fig. 1*B*). In the A^- mutant this sequence was replaced by the highly dissimilar sequence TGCTCCTC and in the T^+ mutant it was replaced by the canonical TATA-box sequence TATATAAT. The binding specificity of these mutants was then tested by gel-mobility shift analysis (Fig. 4). While −37 to −12 fragments containing either the wild-type or T^+ sequence were readily bound by nuclear factors (lanes 2 and 7), no binding was detected with the fragment containing the A^- sequence (lane 12). This latter result confirms that the A_5 segment is part of the binding site recognition sequence. Competition experiments demonstrated that the factor binding to the wild-type sequence is different from the conventional TATA-box binding factor. A 50-fold molar excess of unlabeled wild-type sequence efficiently competed for binding to the wild-type fragment but had no effect on binding to the T^+ fragment (lanes 3 and 9). Similarly, a 50-fold molar excess of unlabeled T^+ sequence efficiently competed for binding to the T^+ fragment but had no effect on binding to the wild-type fragment (lanes 8 and 5). As expected, excess unlabeled A^-

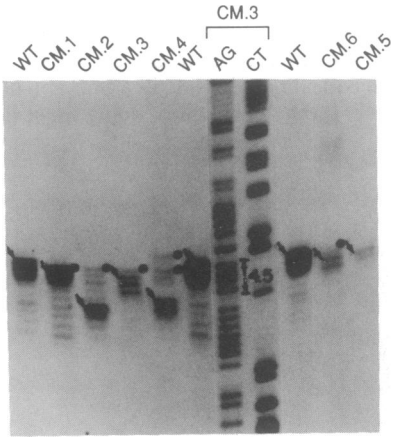

FIG. 3. High-resolution S1 protection analysis of the experiment shown in Fig. 2*B*. A Maxam–Gilbert sequence ladder of an anti strand segment encompassing the cap site of mutant CM.3 is included to illustrate the precise 4.5-nucleotide difference between the cap site position and the largest fragment protected by wild-type transcripts. Mutant CM.3 was used for the sequence ladder because the cap site can be easily recognized as the sole pyrimidine in the anti strand purine tract. Arrows at the left of each lane indicate the major start sites; dots at the right of each lane indicate minor start sites, determined as explained in footnote † of Table 1.

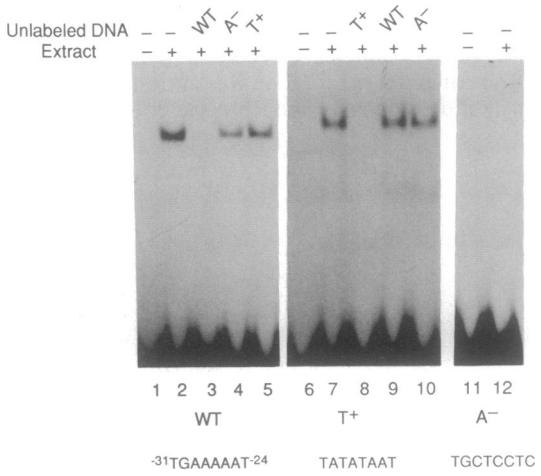

FIG. 4. Characterization of the nuclear factor binding site in the rpS16 TATA-box region. Labeled −37 to −12 fragments containing the indicated wild type (WT) or mutant (T^+, A^-) sequences were incubated with (+) or without (−) nuclear extract from S194 mouse plasmacytoma cells, electrophoresed on 5% polyacrylamide gels, and visualized by autoradiography. As indicated above the lanes, some of the binding reaction mixtures contained a 50-fold molar excess of unlabeled DNA competitor.

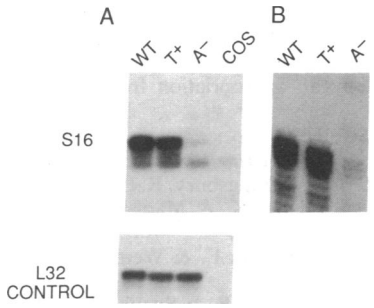

FIG. 5. Expression of rpS16 genes containing different sequences in the TATA-box region. S1-nuclease-protected fragments of a wild-type probe were analyzed on conventional (*A*) and high-resolution (*B*) gels. The shorter (≈60-nucleotide) fragments seen in *A* are due to cross-hybridization with COS cell RNA, which was exceptionally high in this particular experiment.

sequence did not compete significantly for binding to either the wild-type or T^+ fragments (lanes 4 and 10).

When tested by transient transcription assays, the level of expression of the T^+ mutant was indistinguishable from that of the wild-type gene, whereas the expression of the A^- mutant was an order of magnitude lower (Fig. 5*A*). Thus, there is good correspondence between the factor binding capability of the −31 to −24 segment and the efficiency of the rpS16 promoter. Interestingly, the wild-type A_5 element and the canonical TATA box are functionally equivalent despite their different factor binding preferences. This equivalence also pertains to the accuracy of start-site selection (Fig. 5*B*). Indeed, the normal cap site is selected even by the inefficient A^- mutant, indicating that start-site selection is not strongly influenced by interactions in the TATA-box region.

Relationship Between Initiator and TATA-Region Interactions. Further evidence for the apparent independence of interactions in the −31 to −24 region and the initiator region was provided by experiments in which we tested whether the aberrant expression of several of the CM mutants would be altered when the canonical TATA box was substituted for the wild-type A_5 segment. Experiments with a set of double mutants indicated that neither the inaccurate start-site selection nor the relative efficiencies of mutants CM.2, CM.3, CM.4, and CM.6 were affected by the T^+ replacement (Fig. 6, Table 1). High-resolution S1 analysis of these various RNAs confirmed that the cap site of each of the T^+ CM mutants was identical to that of the corresponding CM mutant (data not shown). The fact that the T^+ substitution does not influence the outcome of any of the CM mutations suggests that the TATA-box counterpart and the initiator element make independent contributions to the overall function of the rpS16 promoter.

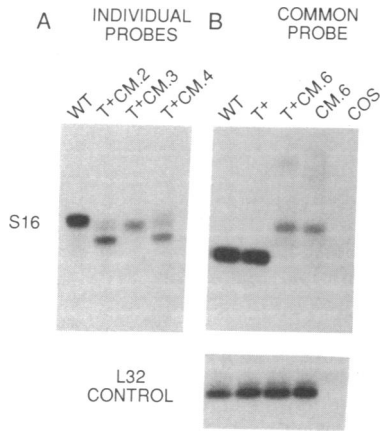

FIG. 6. Expression of rpS16 double mutants containing a canonical TATA box in the −31 to −24 region and selected initiator region mutations. S1 nuclease protection assays with individual probes (*A*) and a common CM.6 probe (*B*), carried out as described for Fig. 2.

DISCUSSION

The Polypyrimidine Initiator. Our *in vivo* mutational analysis of the rpS16 initiator region (−4 to +8) indicates that this element plays an important role in determining the cap site position and that it can also influence the efficiency of RNA polymerase loading. Single purine substitutions at positions −2 or +1 or +3 and a triple purine substitution at −2, +2, and +5 decreased the precision of start-site selection and displaced the site from its normal position by 1–6 bp. Although these mutations did not affect the promoter efficiency, other initiator mutants containing 6, 8, or 12 purine substitutions exhibited a 5-fold decrease in transcript yield as well as a shift in start-site position. Parallel studies of these mutants in a cell-free transcription system (S. Chung and R.P.P., unpublished results) have also indicated that the promoter efficiencies of extensive substitution mutants are lower than the efficiencies of comparable single and triple mutants. Thus, the polypyrimidine initiator sequence critically defines the position of the transcriptional initiation complex, whereas a much less specific sequence is sufficient to satisfy the efficiency requirement. Clearly, an uninterrupted pyrimidine stretch is not necessary for efficient polymerase loading of this promoter.

Two of the single substitution mutants, CM.2 and CM.3, created an initiator element that conforms to the consensus sequence $Y_nCAY_{n'}$, which encompasses the adenosine cap site of several TATA-box-containing genes (16) and at least one TATA-lacking gene (7). In these genes, the A residue within the pyrimidine cluster is invariably selected for transcriptional initiation, whereas in the CM.2 and CM.3 mutants it is avoided in favor of neighboring pyrimidine residues. This avoidance occurs irrespective of whether the rpS16 promoter is functioning with its normal A_5 upstream element or with a substituted TATA box. It would appear, therefore, that different genes can utilize similar initiator sequence information in different ways, depending on the nature of interactions with other gene-specific promoter elements. Indeed, a single A → T mutation at the cap site of the adenovirus major late promoter, which converts the wild-type sequence Y_4CAY_9 to a polypyrimidine sequence Y_4CTY_9, decreases the promoter efficiency by at least 80% (4). In other promoters, mutations in the initiator sequence usually cause some dislocation and imprecision of start site selection (2, 6, 7, 17). As we have observed for rpS16, there does not seem to be any obvious relationship between the sensitivity of these various promoters to initiator mutations and the presence or absence of a canonical TATA-box element.

The TATA-Box Counterpart. None of the mammalian rp promoters examined to date have well-defined TATA boxes in the −20 to −30 region (9, 18, 19). The rpS16 promoter has an AAAAAT sequence at this location, which could conceivably serve as a degenerate TATA element. Yet, our present results clearly indicate that this sequence has a factor binding specificity that is distinct from that of a conventional TATA box. Since a mutation that eliminates the factor binding capacity of the A_5 sequence also greatly reduces the promoter activity, we consider this element to be a functional counterpart of the TATA box. Indeed, when this element is replaced by a more typical TATA box, the promoter activity remains unchanged, although it now interacts with a different factor. We presume that this TATA-specific factor is TFIID, although additional studies with purified extract components will be required to conclusively establish its identity. In any event, our results support the idea that diverse factors

binding to elements in the TATA-box region can play analogous roles in the formation of the transcriptional initiation complex (20, 21).

Mutations that alter the factor binding specificity of the rpS16 TATA-box counterpart have no discernible influence on the positioning of the start site. For rpS16, this positioning function is primarily determined by the initiator element, as is the case for other TATA-lacking promoters that exhibit precise start site selection (6, 7). In such promoters the TATA counterpart serves to increase promoter efficiency in essentially the same manner as other upstream regulatory elements.

Organization of the rpS16 Promoter. The present results, together with previous analyses (10), indicate the following basic structure of the rpS16 promoter (Fig. 1*A*): (*i*) an initiator element that determines the position of the transcriptional start site, but which, by itself, is unable to form an effective initiation complex; (*ii*) three additional elements, B at −83 to −59, C at −20 to −30 (TATA counterpart), and D at +9 to +29, the combined activity of which is essential for rpS16 promoter function; and (*iii*) an Sp1 element at −170 to −161, which is not essential for promoter function, but which increases promoter efficiency by about 2.5-fold. As we have noted previously (10), this basic architecture is similar to that of other well-studied mouse rp promoters, except that the array of essential elements in rpS16 is shifted about 50 bp upstream compared to the others. In many ways, the rp promoter organization is remarkably similar to that of the simian virus 40 late promoter (6), although the sequences of the individual modules are quite different. The great diversity of promoter design, which is becoming evident as more RNA polymerase II promoters are being carefully dissected, suggests that the requirements for building a functional initiation complex may be much more flexible than was previously appreciated.

We are indebted to Drs. John Burch and Tom Kaedesch for critically reviewing this manuscript for us. This research was supported by grants from the National Science Foundation (DCB84-13609) and the National Institutes of Health (AI-17330, CA-06927, RR-05539) and by an appropriation from the Commonwealth of Pennsylvania.

1. Wasylyk, B. (1988) *Crit. Rev. Biochem.* **23,** 77–120.
2. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217,** 316–324.
3. Myers, R. M., Tilly, K. & Maniatis, T. (1986) *Science* **232,** 613–618.
4. Lee, R. F., Cocino, M. F. & Weinmann, R. (1988) *Virology* **165,** 51–56.
5. Tokunaga, K., Hirose, S. & Suzuki, Y. (1984) *Nucleic Acids Res.* **12,** 1543–1558.
6. Ayer, D. E. & Dynan, W. S. (1988) *Mol. Cell. Biol.* **8,** 2021–2033.
7. Smale, S. T. & Baltimore, D. (1989) *Cell* **57,** 103–113.
8. Reynolds, G. A., Goldstein, J. L. & Brown, M. S. (1985) *J. Biol. Chem.* **260,** 10369–10377.
9. Hariharan, N., Kelley, D. & Perry, R. P. (1989) *Genes Dev.* **3,** 1789–1800.
10. Hariharan, N. & Perry, R. P. (1989) *Nucleic Acids Res.* **17,** 5323–5337.
11. Chung, S. & Perry, R. P. (1989) *Mol. Cell. Biol.* **9,** 2075–2082.
12. Siebenlist, U. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 122–126.
13. Wagner, M. & Perry, R. P. (1985) *Mol. Cell. Biol.* **5,** 3560–3576.
14. Kelley, D. E., Coleclough, C. & Perry, R. P. (1982) *Cell* **29,** 681–689.
15. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5718–5722.
16. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209,** 1405–1414.
17. Jones, K. A., Luciw, P. A. & Duchange, N. (1988) *Genes Dev.* **2,** 1101–1114.
18. Chen, I.-T. & Roufa, D. J. (1988) *Gene* **70,** 107–116.
19. Meyuhas, O. & Klein, A. (1989) *J. Biol. Chem.*, in press.
20. Chen, W. & Struhl, K. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2691–2695.
21. Simon, M. C., Fisch, T. M., Benecke, B. J., Nevins, J. R. & Heintz, N. (1988) *Cell* **52,** 723–729.