Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9509–9513, December 1990
Biochemistry

Mutational analysis of an archaebacterial promoter: Essential role of a TATA box for transcription efficiency and start-site selection *in vitro*

(*Sulfolobus*/ribosomal RNA genes/deletion mutants/linker substitution/promoter evolution)

WOLF-DIETER REITER*, UWE HÜDEPOHL, AND WOLFRAM ZILLIG†

Max-Planck-Institut für Biochemie, D-8033 Martinsried, Federal Republic of Germany

Communicated by Walther Stoeckenius, June 11, 1990 (received for review April 3, 1990)

ABSTRACT **By using a recently developed *in vitro* transcription assay, the 16S/23S rRNA-encoding DNA promoter from the archaebacterium *Sulfolobus* sp. B12 was dissected by deletion and linker substitution mutagenesis. The analysis of 5′ and 3′ deletion mutants defined a core promoter region between positions −38 and −2 containing all information for efficient and specific transcription. Further characterization of this region by linker substitution mutagenesis indicated two sequence elements important for promoter function—one located between positions −38 and −25 (distal promoter element) and the other one located between positions −11 and −2 (proximal promoter element). The distal promoter element encompassed the TATA-like "box A" element located approximately 26 nucleotides upstream of the majority of transcription start sites in archaebacteria (Archaeobacteria). All mutations within this box A motif virtually abolished promoter function. Complete inactivation of the proximal promoter element was dependent on extensive mutagenesis; this element is not conserved between archaebacterial promoters except for a high A+T content in stable RNA gene promoters from *Sulfolobus*. Mutants containing insertions or deletions between the distal and proximal promoter elements were only slightly affected in their transcription efficiency but displayed a shift in their major initiation site, retaining an essentially fixed distance between the distal promoter element and the transcription start site. Thus, efficient transcription and start-site selection were dependent on a conserved TATA-like sequence centered approximately 26 nucleotides upstream of the initiation site, a situation unlike that of eubacterial promoters but resembling the core structure of most eukaryotic RNA polymerase II (and some RNA polymerase III) promoters. This finding suggests a common evolutionary origin of these promoters consistent with the known similarities between archaebacterial and eukaryotic RNA polymerases.**

Based on molecular data, archaebacteria (Archaeobacteria) comprise a group of prokaryotic microorganisms phylogenetically distinct from eubacteria and eukaryotes (1–3). Many archaebacteria are characterized by extreme habitats believed to resemble the environmental conditions during the early evolution of life, leading to the suggestion that they represent an ancient group of organisms (1, 2).

Transcription in archaebacteria has primarily been investigated on the level of DNA-dependent RNA polymerases. These studies indicate a single transcribing enzyme displaying a subunit complexity similar to the three eukaryotic nuclear enzymes (4). The characterization of archaebacterial transcription signals has been confined to nucleotide sequence comparisons, revealing two conserved sequence elements, a "box A" motif [consensus TTTA(A or T)A] centered approximately 26 nucleotides upstream of the transcription start site and a "box B" motif [consensus (A or T)TG(A or C)] containing the transcription start site on the central guanosine residue or on another purine nucleotide mapping nearby (5–7). Functional data on the structure of archaebacterial promoters have not been published so far.

Using a recently developed *in vitro* transcription assay (30), we now have dissected the promoter for the single-copy 16S/23S rRNA gene cluster of the extremely thermophilic archaebacterium *Sulfolobus* sp. B12 by extensive deletion and linker substitution mutagenesis. This promoter is particularly well suited for a mutational analysis because it conforms perfectly to the archaebacterial promoter consensus sequence and directs very strong transcription from a single site *in vivo* (8). Our results indicate an essential role of the box A motif for transcription efficiency and start-site selection *in vitro*. In addition, we identified a second promoter element that was not apparent from sequence comparisons.

MATERIALS AND METHODS

Construction of Deletion Mutants. The 5′ deletion mutants 5′Δ-950("pRRN18"), 5′Δ-354, 5′Δ-190, and 5′Δ-125("pRRN-1820"; see Fig. 1) were constructed by cloning the following promoter-containing restriction fragments of *Sulfolobus* 16S/23S rRNA-encoding DNA (rDNA) into pUC18 cleaved with *Sma* I/*Hin*dIII, *Eco*RI/*Hin*dIII, *Acc* I/*Hin*dIII, and *Sma* I/*Hin*dIII, respectively: *Nco* I–*Hin*dIII (5′Δ-950), *Eco*RI–*Hin*dIII (5′Δ-354), *Taq* I–*Hin*dIII (5′Δ-190), and *Dra* I–*Hin*dIII (5′Δ-125). The deletion numbers of these constructs indicate the map position of the first nucleotide of wild-type sequence relative to the transcription start site at +1. Further 5′ deletions were generated by BAL-31 digestion of *Eco*RI-cleaved pRRN1820, subsequent addition of the *Bam*HI linker CGGGATCCCG, cleavage by *Bam*HI and *Hin*dIII, and cloning of the partially deleted fragments into pUC18. 3′ deletions were constructed in an analogous way: we started from *Hin*dIII-cleaved pRRN1820 and used *Bam*HI and *Eco*RI to excise the partially deleted fragments for recloning into pUC18.

Construction of Linker Substitution Mutants. Linker substitution mutants of the pLS series were assembled from matching 5′ and 3′ deletion plasmids essentially as described by McKnight and Kingsbury (9). In short, both types of deletion plasmids were cleaved with the single-cutting enzymes *Bam*HI and *Sca* I, and the large fragment from a 3′ deletion plasmid was ligated to the small fragment from a 5′ deletion plasmid. To obtain mutants with smaller sequence changes, auxiliary

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.

Abbreviations: PPE, proximal promoter element; DPE, distal promoter element; rDNA, rRNA-encoding DNA.

*Present address: Michigan State University–Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, MI 48824.
†To whom reprint requests should be addressed.

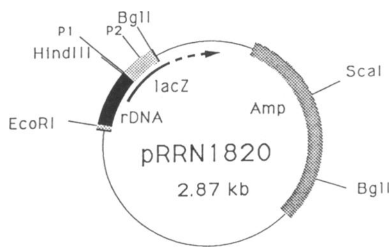

FIG. 1. Physical map of plasmid pRRN1820, representing the parental construct from which all 3′ deletion and linker substitution mutants were generated. The insert DNA consisting of the *Sulfolobus* B12 16S/23S rDNA transcribed leader and upstream sequences is represented by a black bar. An arrow indicates the 5′ end and the direction of transcription for correctly initiated *in vitro* transcripts extending into the *lacZ* region of the vector (stippled bar). Primers used for the synthesis of DNA probes (P1 and P2) and relevant restriction sites are indicated.

*Bam*HI linker substitution mutants were assembled from deletion plasmids differing by two nucleotides in their deletion end points. The central four nucleotides of the linkers were subsequently removed by *Bam*HI cleavage followed by S1 nuclease digestion. Religation of this DNA generated an *Eag* I recognition site. The whole procedure was repeated with *Eag* I for the initial cleavage, giving rise to the pCG series of constructs retaining two nucleotides from the original linker sequence. Mutants containing insertions or deletions within the promoter region (representing the pINS and the pDEL series of plasmids) were constructed similarly, except that the auxiliary *Bam*HI linker substitution mutants were assembled from deletion plasmids differing by variable numbers in their deletion endpoints. The nucleotide sequence of all mutants was verified by dideoxy sequencing (10, 11).

***In Vitro* Transcription Reactions.** *Bgl* I-cleaved plasmid DNA (150 fmol, ≈0.3 μg) was incubated for 10 min at 60°C in 50 μl of a reaction mixture containing 50 mM Tris·HCl (pH 8.0), 25 mM $MgCl_2$, 1 mM EDTA, 1 mM dithiothreitol, 2 mM ATP, 1 mM GTP, 1 mM CTP, 1 mM UTP, and 8 μl of a cell-free extract prepared from late-logarithmic-growth *Sulfolobus* B12 (30). After the mixture was chilled on ice, 50 μl of 50 mM EDTA and 15 μg of *Escherichia coli* tRNA were added, and the reaction mixture was extracted three times with 100 μl of phenol/chloroform/isoamyl alcohol, 24:24:1 (vol/vol). After ethanol precipitation of 70 μl of the aqueous phase, the template DNA was removed by a 30-min incubation at room temperature with 23 units of RNase-free DNase (Boehringer Mannheim) in 50 μl of the reaction buffer recommended by the manufacturer. After addition of 50 μl of 50 mM EDTA, the reaction mixture was extracted twice with phenol/chloroform/isoamyl alcohol and once with chloroform/isoamyl alcohol, 24:1 (vol/vol). The aqueous phase was stored at −70°C.

Preparation of DNA Probes. DNA probes for S1 nuclease analysis were prepared by extension of 5′-^{32}P-labeled primers hybridized to single-stranded DNAs containing wild-type or mutant promoter sequences transferred from pUC18 derivatives into M13mp18. For the analysis of 5′ deletion derivatives and linker substitution mutants containing wild-type sequences downstream of position −8, the DNA probe was prepared from an M13mp18 derivative containing the insert of pRRN18. In all other cases, mutant-specific DNA probes were prepared from M13mp18 derivatives containing the respective inserts. In the case of 3′ deletion mutants, the 17-mer oligonucleotide 5′-TTCGCTATTACGCCAGC-3′ mapping within the *lacZ* region of the vector was used for primer extension (primer P2 in Fig. 1). In all other cases, synthesis of DNA probes was initiated from the phage M13 universal sequencing primer 5′-GTAAAACGACGGCCAGT-3′ (primer P1 in Fig. 1). The conditions for hybridization, primer extension, and purification of DNA probes were as described (12). In experiments where several different DNA probes were required for RNA quantitation, the same labeled oligonucleotide was used for probe synthesis, ensuring identical specific activities.

S1 Nuclease and Primer Extension Analysis of *in Vitro* RNA. Two microliters of *in vitro* RNA prepared as described above were hybridized to a molar excess of at least 5-fold of DNA probe. Conditions for hybridization, S1 nuclease digestion, and electrophoresis through denaturing polyacrylamide gels were as described (12). Transcription efficiencies were quantitated by densitometry of the autoradiographs with a calibration curve to correct for nonlinearity. Primer extension analysis of *in vitro* RNA was carried out essentially as described (8).

RESULTS

Construction and Analysis of Deletion Mutants. As a first step to define promoter elements for the *Sulfolobus* 16S/23S rRNA gene cluster, 5′ and 3′ deletion mutants were constructed as described. With the exception of the 3′ deletion

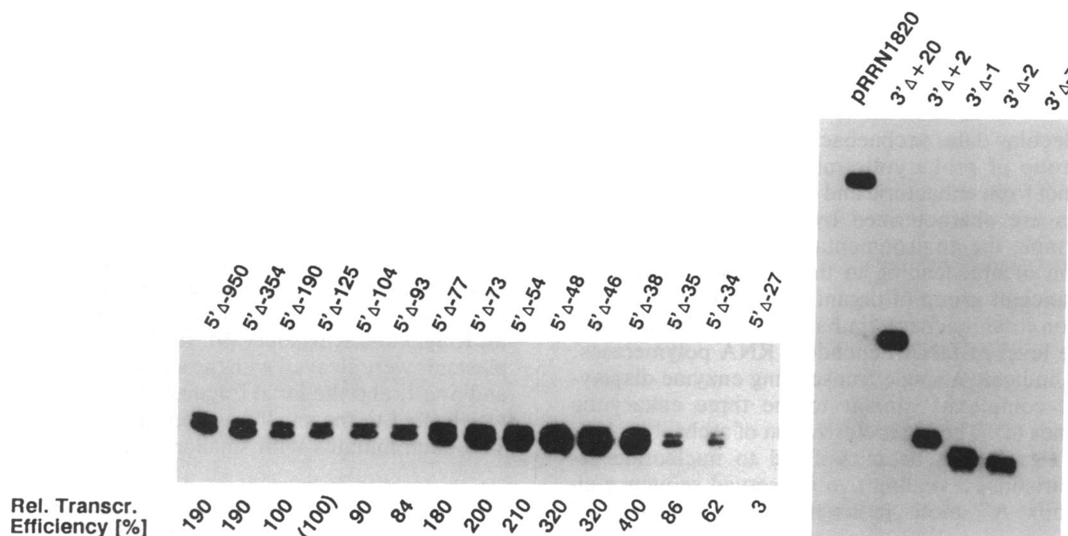

FIG. 2. S1 nuclease analysis of *in vitro* transcripts derived from 5′ and 3′ deletion mutants. The deletion endpoints indicated above each lane correspond to the map position of the first nucleotide of wild-type sequence. (*Left*) 5′ deletions. Transcription efficiencies relative to the −125 deletion (pRRN1820) are indicated below the lanes. (*Right*) 3′ deletions.

mutants, all constructs contained 95 base pairs of the rDNA transcribed leader fused to the pUC18 *lacZ* region. The 3′ deletion mutants were derived from plasmid pRRN1820 (Fig. 1), which contained 125 nucleotides of wild-type *Sulfolobus* B12 DNA upstream from the transcription start site. After incubation with a cell-free extract, the amount of correctly initiated transcript extending into vector sequences was quantified by S1 nuclease analysis with an antisense DNA probe. Starting from an initial construct containing 0.95 kilobases of upstream sequences ("pRRN18"), a deletion to position −354 had no influence on transcription efficiency, whereas a deletion to position −190 resulted in a 50% reduction, indicating the removal of positive regulatory sequence(s) (Fig. 2 *Left*). Further deletions to positions −125, −104, and −93 did not result in significant changes. Removal of additional sequences lead to a steady increase in the amount of *in vitro* RNA (4-fold in the 5′Δ-38 mutant) apparently because of the loss of negative regulatory sequence(s) (Fig. 2 *Left*). Deletions to position −35 and −34 were significantly reduced in their transcription efficiency, and a

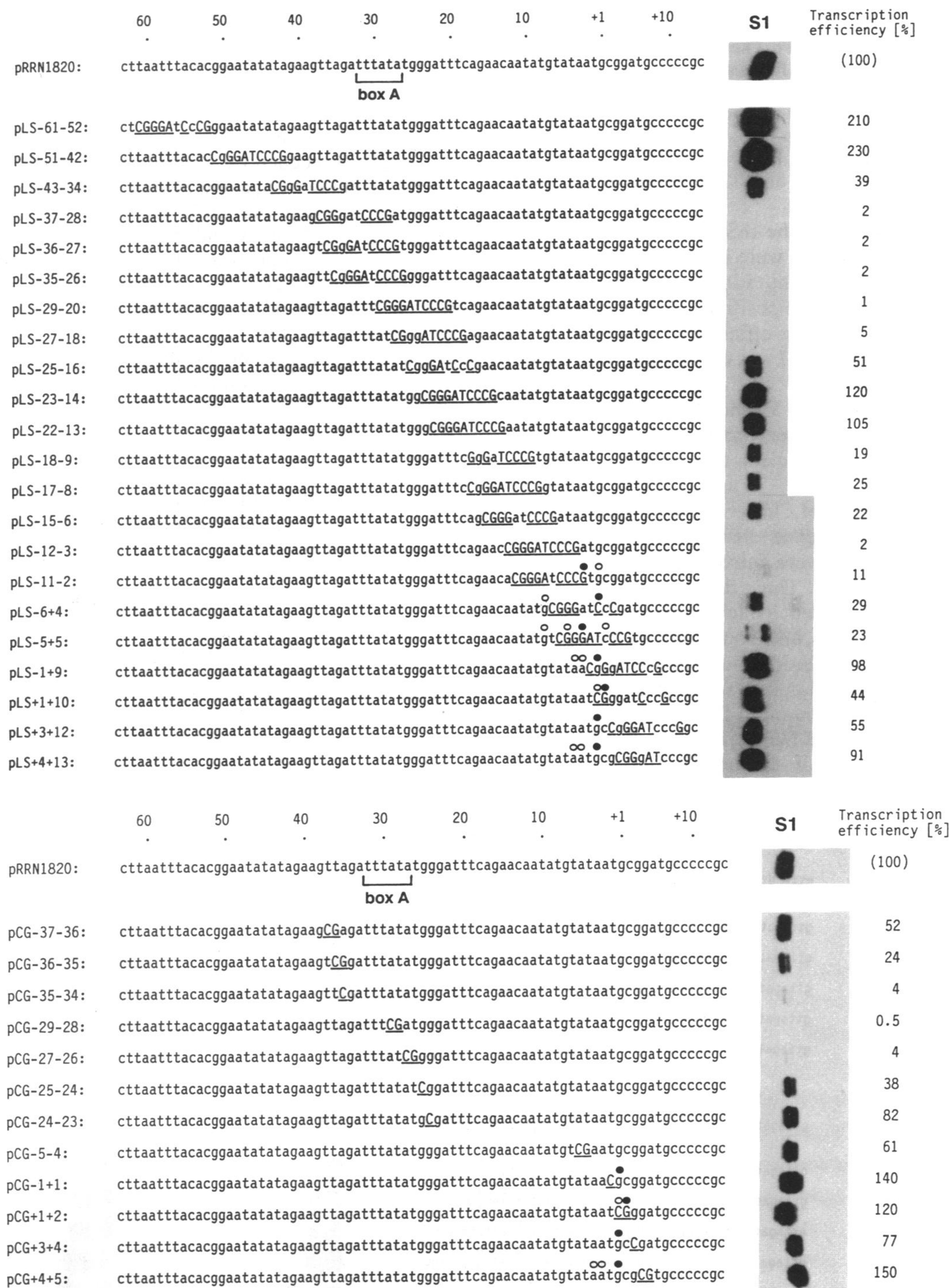

FIG. 3. Linker substitution analysis of the 16S/23S rDNA promoter. Wild-type nucleotides in the template sequences are shown in lowercase letters, and mutant nucleotides are shown in underlined uppercase letters. The box A motif defined by sequence comparisons is indicated. S1 nuclease-generated fragments corresponding to specifically initiated transcripts are shown to the right of the template sequences. Major and minor initiation sites determined by primer extension analysis are indicated by closed and open circles, respectively. (*Upper*) Mutants containing a decanucleotide *Bam*HI linker. (*Lower*) Mutants containing a CG dinucleotide substitution.

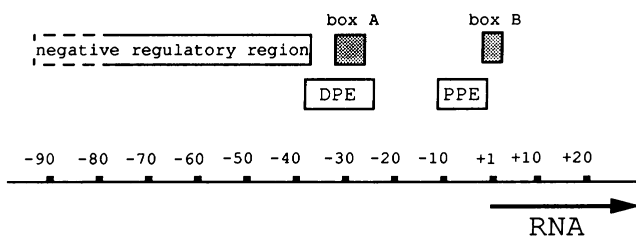

FIG. 4. Schematic representation of regulatory elements of the *Sulfolobus* 16S/23S rDNA promoter. Regions covered by regulatory sequences as determined by mutational analysis are indicated by open bars. Note that the 5′ border of the negative regulatory element has been mapped at low resolution. Presumptive positive regulatory sequences located further upstream are not shown. Sequences conserved between most archaebacterial promoters (boxes A and B) are indicated by stippled rectangles.

deletion to position −27 almost abolished promoter function (Fig. 2 *Left*).

To define the 3′ boundary of the 16S/23S rDNA promoter, a number of 3′ deletion mutants were investigated by *in vitro* transcription and transcript quantitation as above. As shown in Fig. 2 *Right*, 3′ deletions up to position −2 were essentially wild-type in their transcription efficiency. However, a 3′ deletion extending to position −7 was only poorly transcribed (Fig. 2 *Right*).

The combined results from 5′ and 3′ deletion mutants indicate that the core promoter region encompassing sequences indispensable for efficient transcription initiation was located between positions −38 and −2.

Introduction and Analysis of Clustered Point Mutations. To characterize essential sequence elements in more detail, clustered point mutations were introduced by a modified linker substitution approach. Three different classes of mutant templates were generated, which contained mutations confined to segments of 10, 6, and 2 nucleotides, respectively (Fig. 3; data for the 6-nucleotide mutants are not shown). Clustered point mutations mapping immediately upstream of the core promoter region were more than 2-fold increased in their promoter activity (Fig. 3 *Upper*, mutants pLS-61-52 and pLS-51-42), consistent with a similar effect of 5′ deletions mapping between positions −77 and −38 (Fig. 2 *Left*).

The mutational analysis of the core promoter region revealed two sequence elements essential for promoter function—one of them located between positions −38 and −25 (distal promoter element, DPE) and the other one located between positions −11 and −2 (proximal promoter element, PPE; see Figs. 3 and 4). The DPE encompassed the box A consensus region defined by sequence comparisons (positions −32 to −27). All mutations within this conserved motif virtually abolished promoter function (Fig. 3). Complete inactivation of the PPE required extensive sequence changes throughout this element (Fig. 3; compare the mutants pLS-15-6 and pLS-5+5 covering part of the PPE, with mutant pLS-12-3 covering the entire PPE).

Clustered point mutations changing the wild-type start and/or adjacent sequences usually resulted in multiple initiation sites, the major one located close to the wild-type +1 position (Fig. 3). In most cases, initiation occurred on purine nucleotides preceded by a pyrimidine residue.

Transcription of Spacing Mutants. To determine how the distance between the DPE and the PPE may influence transcription efficiency and/or start site selection, a number of mutants were constructed containing insertions or deletions between these two elements. These mutants were transcribed with efficiencies between 25% and 100% of wild-type level (Fig. 5). There was no clear correlation between the lengths of the insertions or deletions and the promoter strength. With the exception of the one-nucleotide deletion mutant (pDEL1), all templates were transcribed from multiple sites invariably representing purine nucleotides preceded by a pyrimidine residue. The major initiation site was shifted to a position essentially retaining the wild-type distance from the DPE. Nonetheless, some initiation at the wild-type +1 position was observed in all of the mutants investigated (Fig. 5).

DISCUSSION

We have dissected an archaebacterial promoter by deletion and linker substitution mutagenesis, using an *in vitro* system to monitor transcription efficiency and start-site selection. The *Sulfolobus* B12 16S/23S rDNA promoter was chosen for this study because it is efficiently transcribed *in vivo* (8) and conforms perfectly to the promoter consensus sequences defined by sequence comparisons (5–7). The analysis of 5′

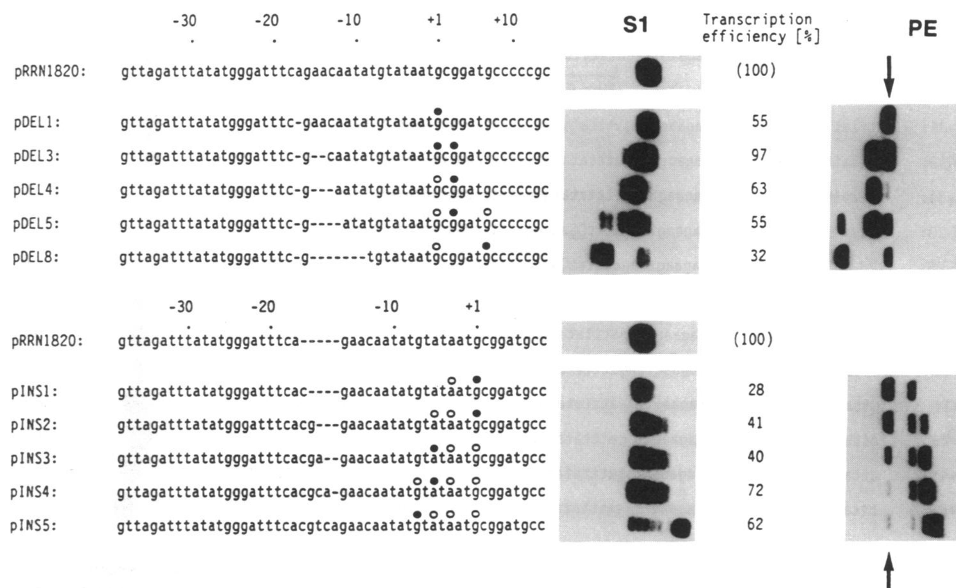

FIG. 5. S1 nuclease and primer-extension (PE) analyses of transcripts derived from spacing mutants. Major and minor initiation sites determined by primer-extension analysis are indicated by closed and open circles, respectively. Primer-extension products corresponding to the wild-type start site are marked by arrows.

deletion mutants indicated a weak positive regulatory region between positions −354 and −190 and negative regulatory sequences between positions −93 and −38. We consider it unlikely that the transcription enhancement observed for the latter group of mutants is due to the proximity of vector sequences since linker substitutions within this region had a similar effect. The function of this negative regulatory region is presently unknown; it may be involved in growth rate-dependent control of stable RNA synthesis.

The analysis of 5′ and 3′ deletions indicated that the core promoter region containing sequences required for efficient initiation of transcription was located between positions −38 and −2. Linker substitution mutagenesis within this region identified two essential sequence elements, a DPE between positions −38 and −25 and a PPE between positions −11 and −2. The sequence TGC encompassing the wild-type start on the central guanosine residue was not required for transcription efficiency but appeared to be involved in unambiguous start-site selection. This sequence essentially corresponds to the box B motif [consensus (A or T)TG(A or C)] conserved between most archaebacterial initiation sites.

The identification of the PPE was unexpected because it was not apparent from sequence comparisons. A reevaluation of published archaebacterial promoter sequences revealed that this region is generally A+T-rich in stable RNA promoters from *Sulfolobus* (8, 13, 14). Since the linkers used for mutagenesis were very G+C-rich, it appears possible that the function of the PPE is dependent on an A+T-rich nucleotide composition rather than on a specific DNA sequence.

The DPE encompassed the box A sequence [consensus TTTA(A or T)A] conserved between almost all archaebacterial promoters characterized so far. Our data clearly indicate that this motif is essential for transcription efficiency and that it represents the primary determinant of start site selection.

Based on these data, the archaebacterial promoter structure appears to be unlike that of eubacterial promoters, which are composed of two defined sequence elements typically centered at positions −35 and −10 (for review, see refs. 15–18). Although it may be tempting to equate the −35 region with the DPE and the −10 region with the PPE, we consider this unjustified for the following reasons. First, there is no sequence similarity between the eubacterial −35 region and the DPE. Second, start-site selection depends on the −10 region in eubacteria and on the DPE in archaebacteria. Furthermore, the mutants with an altered spacing between DPE and PPE were not dramatically reduced in their transcription efficiency, whereas eubacterial promoters are very sensitive to distance changes between the −35 and the −10 region (19). For these reasons the term "prokaryotic promoter," firmly established in the scientific literature, appears inappropriate because it does not apply to archaebacteria.

A comparison of the archaebacterial promoter structure with promoters in eukaryotes indicates intriguing similarities with the core structure of eukaryotic RNA polymerase II promoters (reviewed in refs. 20–23) and the TATA box-containing class of RNA polymerase III promoters (reviewed in refs. 24 and 25). The common feature of these promoters is the presence of an A+T-rich sequence element typically located 25–30 nucleotides upstream of the initiation site. In both archaebacteria and eukaryotes, this sequence element is involved in transcription efficiency and start-site selection. The similarity in promoter structure is paralleled by similarities between archaebacterial and eukaryotic RNA polymerases (particularly polymerases II and III; refs. 4 and 26–29), suggesting that both expression signals and transcribing enzymes have been conserved between these two primary kingdoms.

We wish to thank Gabriela Simon for excellent technical assistance and Felicitas Pfeifer for critical reading of the manuscript.

1. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5088–5090.
2. Woese, C. R., Magrum, L. J. & Fox, G. E. (1978) *J. Mol. Evol.* **11,** 245–252.
3. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
4. Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. (1983) *EMBO J.* **2,** 1291–1294.
5. Wich, G., Hummel, H., Jarsch, M., Bär, U. & Böck, A. (1986) *Nucleic Acids Res.* **14,** 2459–2479.
6. Reiter, W.-D., Palm, P. & Zillig, W. (1988) *Nucleic Acids Res.* **16,** 1–19.
7. Thomm, M. & Wich, G. (1988) *Nucleic Acids Res.* **16,** 151–163.
8. Reiter, W.-D., Palm, P., Voos, W., Kaniecki, J., Grampp, B., Schulz, W. & Zillig, W. (1987) *Nucleic Acids Res.* **15,** 5581–5595.
9. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217,** 316–324.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
11. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
12. Reiter, W.-D. & Palm, P. (1990) *Mol. Gen. Genet.* **221,** 65–71.
13. Kaine, B. P., Gupta, R. & Woese, C. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3309–3312.
14. Kaine, B. P. (1987) *J. Mol. Evol.* **25,** 248–254.
15. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13,** 319–353.
16. McClure, W. R. (1985) *Annu. Rev. Biochem.* **54,** 171–204.
17. Reznikoff, W. S., Siegele, D. A., Cowing, D. W. & Gross, C. A. (1985) *Annu. Rev. Genet.* **19,** 355–387.
18. Ishihama, A. (1988) *Trends Genet.* **4,** 282–286.
19. Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H. & Ikehara, M. (1983) *Nucleic Acids Res.* **11,** 5855–5864.
20. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
21. Bucher, P. & Trifonov, E. N. (1986) *Nucleic Acids Res.* **14,** 10009–10026.
22. Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987) *Science* **236,** 1237–1245.
23. Wasylyk, B. (1988) *CRC Crit. Rev. Biochem.* **23,** 77–120.
24. Sollner-Webb, B. (1988) *Cell* **52,** 153–154.
25. Murphy, S., Moorefield, B. & Pieler, T. (1989) *Trends Genet.* **5,** 122–126.
26. Berghöfer, B., Kröckel, L., Körtner, C., Truss, M., Schallenberg, J. & Klein, A. (1988) *Nucleic Acids Res.* **16,** 8113–8128.
27. Leffers, H., Gropp, F., Lottspeich, F., Zillig, W. & Garrett, R. A. (1989) *J. Mol. Biol.* **206,** 1–17.
28. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4569–4573.
29. Pühler, G., Lottspeich, F. & Zillig, W. (1989) *Nucleic Acids Res.* **17,** 4517–4534.
30. Hüdepohl, U., Reiter, W.-D. & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5851–5855.