Volume 16 Number 1 1988 Nucleic Acids Research

An archaebacterial promoter element for stable RNA genes with homology to the TATA box of higher eukaryotes

Michael Thomm and Günter Wich[1]

Lehrstuhl für Mikrobiologie, Universität Regensburg, Universitätsstr. 31, D-8400 Regensburg and [1]Lehrstuhl für Mikrobiologie, Universität München, Maria-Ward-Str. 1a, D-8000 München 19, FRG

Received September 23, 1987; Revised and Accepted December 3, 1987

ABSTRACT

The RNA polymerase of Methanococcus vannielii, in binary complex with two stable RNA operons, protects from exonuclease digestion the region from 32 bp upstream (-32) to 18 bp downstream (+18) of the transcription start site. Contained within this binding region, centered at -25, is an AT-rich sequence which is highly conserved upstream of 26 other archaebacterial tRNA and rRNA genes. We therefore propose the sequence TTTA$^{T}_{A}$ATA as a common element of promoters for stable RNA genes in archaebacteria. Both the similarity in sequence and the location of this conserved octanucleotide suggest homology to the eukaryotic TATA box preceding protein encoding genes transcribed by RNA polymerase B.

INTRODUCTION

There are fundamental differences in the mechanisms by which genes are transcribed in eukaryotic and in prokaryotic cells. Eubacteria - which constitute one of the two prokaryotic lineages (1) - possess a single RNA polymerase core enzyme. The eubacterial promoter is often distinguished by two primary structural elements, located 10 bp and 35 bp upstream of the site where transcription starts (2-4).

Eukaryotes, on the other hand, exhibit three different types of RNA polymerases which do not share any common recognition sequences at the DNA level (5-7). In promoters for RNA polymerase B, however, the consensus DNA sequence 5'TATA$^{T}_{A}$A$^{T}_{A}$ 3' (the TATA box) exists 25 bp upstream of the start nucleotide of mRNA (5).

The discovery of archaebacteria as a second prokaryotic line of descent (1,8) and the structural similarities of their RNA polymerase with the eukaryotic transcription enzymes (9) poses the question on the structure of archaebacterial promoters. In

studies performed on several archaebacteria, possible promoter elements were suggested from comparative sequence analyses (10-13). However, no archaebacterial consensus promoter sequence was detected (14). Up to now it has been unclear which segments of these sequences are the recognition signals for the RNA polymerase. As an approach to investigate promoter structures in archaebacteria, we studied binding of the RNA polymerase of the methanogen Methanococcus vannielii to DNA fragments containing the upstream regions of its rRNA and tRNA operons (12).

MATERIALS AND METHODS

Polyacrylamide gel electrophoresis of polypeptides.
Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS) was performed as indicated previously (15).

Labelling, isolation and analysis of DNA fragments.
The methods used for labelling/purifying and DNA sequence analysis of DNA fragments are indicated in reference (12).

Purification of the RNA polymerase of Methanococcus vannielii.
The enzyme was enriched analogous to reference (15). Further purification was attained by fast liquid pressure chromatography using a molecular sieve (TSK-3000, LKB) and, subsequently, and anion exchange (Mono Q, Pharmacia) column. The yield was about 22 units/g wet cells. One unit (U) of activity is that amount of enzyme which incorporates 1 nmol AMP in 10 min at 40°C under the conditions described previously (15).

Preparation of probe DNA's.
The recombinant plasmids pMV1 and pMT21 (12), were used to isolate DNA fragments harbouring the upstream region of an rRNA and tRNA operon of M. vannielii. To identify the 5' end of the binding side, a 387 bp NcoI/EcoRV fragment of pMV1 (see Fig. 1a) and a 248 bp DraI/AvaI (see Fig. 1c) fragment of pMT21 labelled at the right (intergenic) end were purified. For the location of the 3' end of the promoter, different fragments labelled at the opposite side had to be isolated. A 422 bp TaqI/HpaII fragment of pMV1 and a 303 bp DdeI/BamHI fragment of

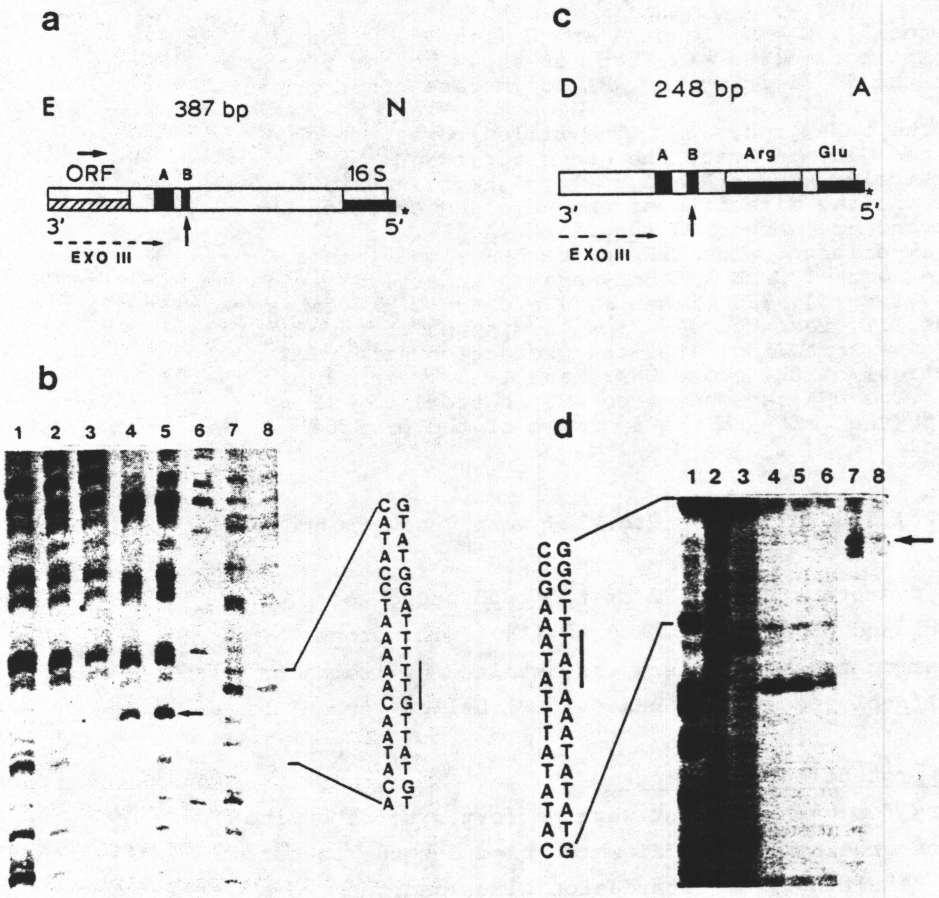


Figure 1. Location of the 5' boundaries of the M. vannielii rRNA and tRNA promoter.
a) Schematic representation of the rRNA probe DNA harbouring an open reading frame (ORF; hatched box), the two sequences conserved between stable RNA genes of M. vannielii (box A and box B (12)), the start site of transcription (↑) (12) and 70 bp of the DNA coding for mature 16S (filled box). The asterisk indicates the labelled (codogenic) DNA strand. The direction of degradation of the labelled DNA strand by exonuclease III and of transcription of the ORF are shown (- - - - →; →).
E (EcoRI). N (NcoI).
b) Protection experiments. Analyses of fragments generated from the rRNA probe DNA by exonuclease III action in the absence (lanes 1-3) and presence of 132 (lane 4) and 66 mU (lane 5) of RNA polymerase. The exonuclease concentrations were 1500 (lanes 3,4) 750 (lanes 2,5) and 400 U/ml (lane 1), respectively, the incubation time 15 min. The arrow indicates the protected DNA fragment. DNA sequencing reactions of the

probe-DNA: G (lane 6), C + T (lane 7) and C (lane 8). The DNA sequence from -25 (bottom) to -43 (top) is shown to the right. The bar denotes the 5' limit of the RNA polymerase binding site.
c) Diagram of the tRNA probe DNA. The filled boxes, designated Arg and Glu, indicate the genes for $tRNA^{Arg}$ and $tRNA^{Glu}$, respectively. Boxes A and B (12), the transcription start site (12) and the direction of exonuclease action on the labelled DNA strand are shown. D (DdeI); A (AvaI).
d) Protection experiment. The tRNA probe DNA alone (lanes 4-6) and in complex with 132 mU RNA polymerase (lanes 7,8), was exposed to 1500 (lane 4), 750 (lanes 5,7) and 400 U/ml (lanes 6,8) exonuclease III, respectively, for 10 min and the products were analyzed. The arrow indicates the protection site. DNA sequencing reactions of the probe DNA: lane 1 (G), lane 2 (C+T) and lane 3 (C). The DNA sequence from -23 (bottom) to -39 (top) is shown to the left and the 5' limit of the promoter marked by the bar.

pMV1 (see Fig. 2a) and pMT21 (see Fig. 2c) were used for these experiments.

The eubacterial promoters P1 and P2 of the rrnB operon of E. coli have been cloned on plasmid pKK65-10 (16). The probe DNA (Fig. 3a) harbouring these promoters was obtained by isolation of a 492 bp SspI/EcoI fragment of pKK65-10, labelled within the 16S rRNA gene.

Exonuclease III protection experiments.

Pilot experiments for each fragment were performed to determine the concentration of exonuclease III which produces the lowest number of bands resulting from retardation sites and the highest yield of "half-way" (17) DNA molecules (fragments degraded approximately to the midpoint). The exonuclease protection experiment was carried out in 100 µl of a mixture containing 5 pmoles (rRNA gene) and 6 pmoles (tRNA gene) end-labelled DNA in 20 mM Tris/HCl (pH 8.0), 10 mM $MgCl_2$, 50 mM KCl, 0.1 mM EDTA, 250 µg bovine serum albumin/ml and the amounts of RNA polymerase as indicated. In control experiments, purification buffer (15) was added instead of RNA polymerase. The samples were incubated for 20 min at 30°C then, after addition of exonuclease III, for a further 10 or 15 min, as indicated. The nuclease reaction was stopped by adjusting the samples to 0.3 M ammonium acetate, 50 mM EDTA, and 10 µg/ml calf thymus DNA. The samples were extracted with an equal volume of phenol/chloroform (1:1). DNA was precipitated

with ethanol and the fragments produced by the exonuclease treatment analyzed on a 6 percent polyacrylamide-urea sequencing gel.

RESULTS

Mapping of the 5' limits of the M. vannielii rRNA and tRNA promoter.

To locate the binding site on promoters for stable RNA's, the binary complex of RNA polymerase and DNA-molecules labelled with ^{32}P at a unique 5' end was exposed to exonuclease III. Proteins firmly bound to specific DNA sequences block the movement of the 3' 5' exonuclease (4,17). The area protected from digestion by the RNA polymerase molecule can therefore be deduced from determination of the first points on either side at which exonucleolytic degradation of the DNA is stopped (4). To detect the left (5') end of the promoter, we labelled DNA fragments on the codogenic DNA strand within the 16S rRNA (Fig. 1a) and $tRNA^{Arg}$-gene (Fig. 1c). These were isolated and digested with exonuclease III in the presence and absence of RNA polymerase. The degradation products were then compared electrophoretically in urea-polyacrylamide gels. In experiments where RNA polymerase was omitted, a distinct pattern of bands was noted (Fig. 1b, lanes 1-3; Fig. 1d, lanes 4-6). This implies the existence of retardation sites of the exonuclease, which have been previously shown to occur with DNA fragments (17,18). However, when the RNA polymerase was added, one strong additional band appeared in each case (Fig. 1b, lanes 4,5; Fig. 1d, lanes 7,8). In a series of control experiments using various exonuclease concentrations and incubation times, these bands could never be detected in the absence of RNA polymerase (data not shown). The bands originating from pausing of the exonuclease at the retardation sites below this position were reduced in intensity in these lanes. These findings indicate that the bulk of exonuclease molecules was obstructed at specific regions due to the binding of RNA polymerase. The size of the protected DNA fragments were determined by comparison with DNA sequence standards (Fig. 1b, lanes 6-8; Fig. 1d, lanes 1-3). In this way, the 5'

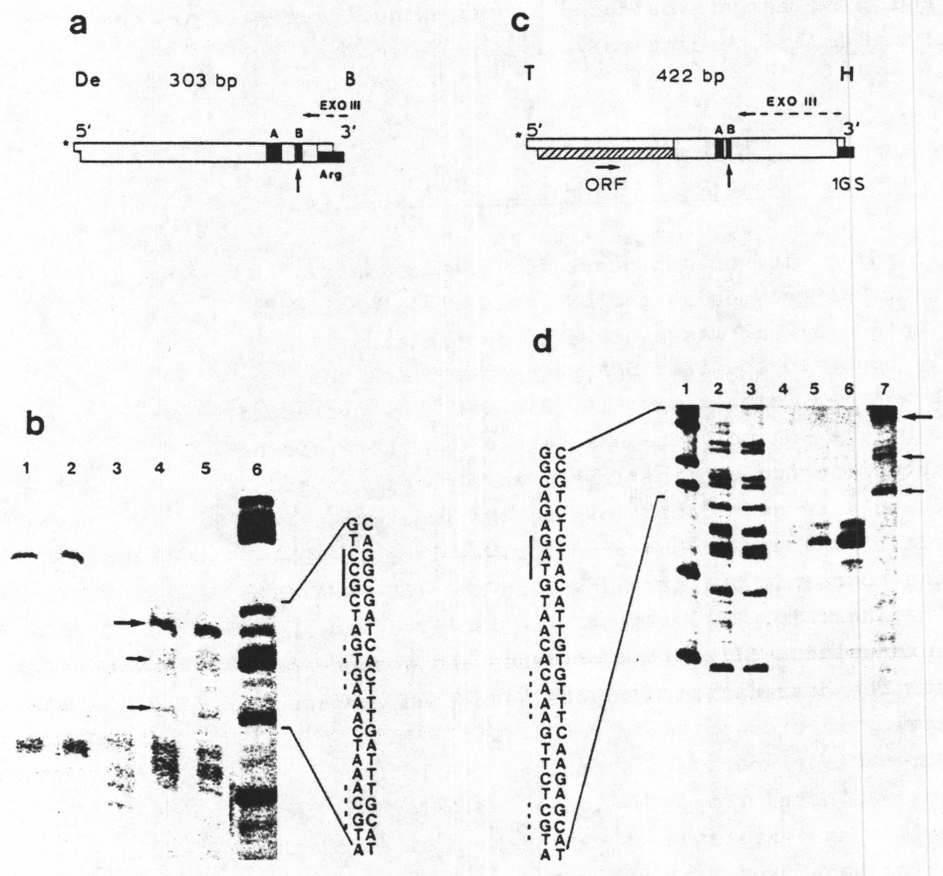


Figure 2. Location of the 3' boundaries of the M. vannielii rRNA and tRNA promoter.
a) Diagram of the rRNA probe DNA labelled at the noncoding DNA strand (asterisk). T (TaqI); H (HpaII). The other abbreviations are as in Fig 1a.
b) Protection experiment. Analysis of the rRNA probe DNA exposed for 10 min to 1500 (lanes 1,4), 750 (lanes 2,5) and 400 U/ml (lane 3) of exonuclease III, respectively, in the absence (lanes 1-3) and presence of 66 mU of RNA polymerase (lanes 4,5). Lane 6, G specific DNA sequencing region of the labelled TaqI/HpaII fragment. The major and minor protection sites are indicated by larger and smaller arrows. The DNA sequence from -3 (bottom) to +20 (top) is show to the right. The bar and dotted lines beside the sequence denote the major and minor protection sites, respectively.
c) Representation of the tRNA probe DNA, labelled at the noncoding DNA strand (asterisk). De (DdeI); B (BamHI). The other abbreviations are as defined in Fig. 1c.
d) Protection experiment. Analyses of control digestions of the DdeI/BamHI restriction fragment with 1600 (lane 4), 800

(lane 5) and 400 units/ml (lane 6) of exonuclease III, respectively. Lane 7, the same as 5, but in the presence of 132 mU RNA polymerase. The DNA sequence from -2 (bottom) to +26 (top) is shown to the left. Major and minor protection sites are indicated as in Fig. 2b.

boundary of the RNA polymerase binding site on the rRNA operon and on the tRNA operon was found to be 31 (± 1) bp and 32 (± 1) bp, respectively, upstream from the start of transcription.

Mapping of the 3' limits of the rRNA and tRNA promoter.

Similar DNA fragments harbouring the intergenic regions of the same operons were used for exonuclease protection experiments designed to identify the right (3') end of the RNA polymerase binding site. In this case, however, the DNA was labelled at the opposite 5' end on the noncoding DNA strand (Fig. 2a, c). For both genes three bands were observed which were dependent upon the presence of RNA polymerase (Fig. 2b, lanes 4 and 5; Fig. 2d, lane 7). The major protection site corresponding to the predominant band was consistently located 18 bp downstream from the start of transcription. The presence of faint bands, which were located at +8 and +1 respectively, may indicate that the region from +18 to the start of transcription is less well protected by the RNA polymerase.

Does the Methanococcus RNA polymerase recognize a strong eubacterial promoter?

A control experiment on a DNA-fragment harbouring the P1 promoter of the Escherichia coli rrnB operon (19) (Fig. 3a), gave no evidence for protection by the archaebacterial RNA polymerase (Fig. 3b, lanes 7,8) indicating that this enzyme does not recognize this typical eubacterial promoter. On the same promoter, the 5' limit of the E. coli RNA polymerase binding site was located by exonuclease III protection experiments and found to be 43 bp upstream of the transcription site (Fig. 3b, lanes 9,10). The same 5' limit has been reported for a different eubacterial promoter identified by the exonuclease protection technique (4).

On all restriction fragments used in the study in addition to the specific protection site also non-specific protection of the region from the end of a particular fragment to about 30-50

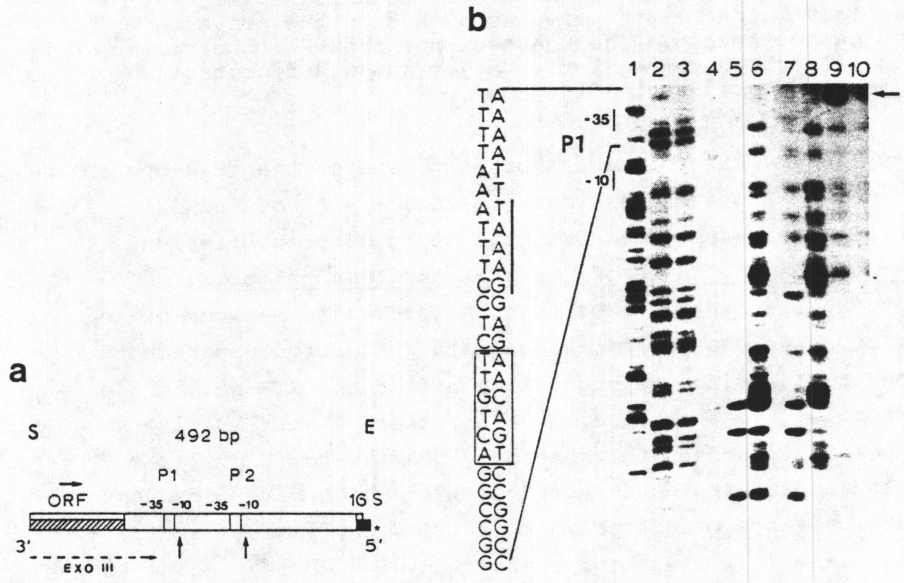


Fig 3. DNA binding analysis of the E. coli and M. vannielii RNA polymerase to a ribosomal RNA promoter of E. coli.
a) Diagram of the DNA fragment harbouring the promoter P1 and P2 of the rrnB operon (19) of E. coli, the two start sites of transcription (16) and the beginning of the 16S rRNA gene (filled box). The other symbols are defined in Fig. 1a.
b) Protection experiments. The eubacterial DNA (5 pmoles) was degraded for 15 min in the absence (lanes 4-6) and in the presence of 66 mU of M. vannielii (lanes 7,8) and 50 mU of E. coli RNA polymerase (lanes 9,10). The exonuclease concentration was 200 (lane 6), 400 (lanes 5, 8, 10) and 800 U/ml (lanes 4, 7, 9). Lanes 1-3 (DNA sequencing reactions of the probe DNA: G (lane 1), C+T (lane 2) and C (lane 3). The sequences of P1 is shown to the left. The bar indicates the 5' limit of the binding site of the E. coli RNA polymerase. Binding to P2 cannot be detected in this experiment as most exonuclease molecules are blocked by E. coli RNA polymerase at P1 (note the reduced intensities of bands caused by retardation sites in lanes 9 and 10).

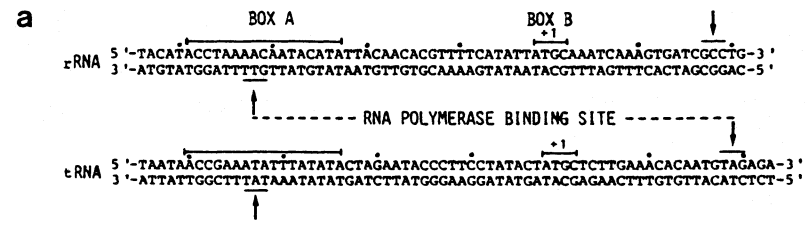
bp downstream was observed with both the Methanococcus and the E.coli RNA polymerase (data not shown). In one case also strong binding to a single stranded nick occurred. This indicates that preferred non-specific binding to free ends of DNA and nicks, which has been reported for eubacterial (20) and eukaryotic (21) RNA polymerases is also a property of the Methanococcus enzyme.

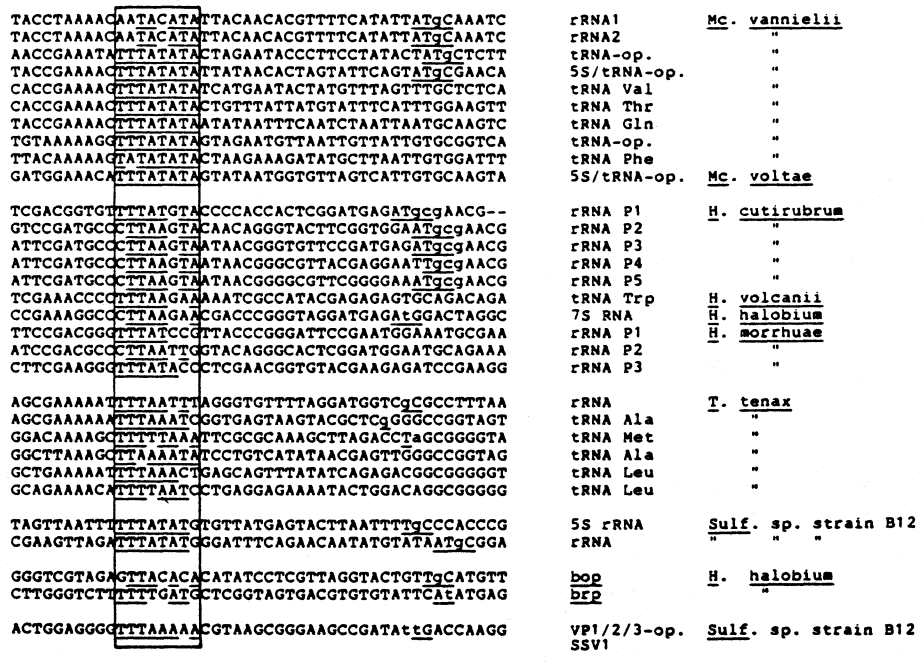
DISCUSSION

Evidence for an archaebacterial consensus promoter sequence

The control experiments and the fact that identical RNA polymerase binding sites were found for two genes of Methanococcus confirm the validity of our approach. From our results we conclude that the enzyme covers the stretch from -32 to +18 on promoters of this organism (Fig. 4a). Thus the nuclease protection studies show that the two conserved sequences noted recently (12) at the start of transcription, box B, and 20 to 40 bp upstream of stable RNA genes of M. vannielii, box A, are indeed involved in RNA polymerase binding. However, only 8 bp of box A, namely TTTATATA, are protected by the RNA polymerase. This finding suggests that this octanucleotide is the essential part of the box A motif. Alignment of the sequences at the 5'end of 26 archaebacterial stable RNA genes revealed that this octanucleotide is the single DNA region which is conserved between all representatives of the methanogenic/halophilic (M. vannielii (12), M. voltae (12), Halobacterium cutirubrum (11), H. volcanii (22), H. halobium (23), Halococcus morrhuae (24)) and the sulphur-dependent (8) branch of archaebacteria (Thermoproteus tenax (13); Fig. 4b, Sulfolobus sp. strain B 12 (25)). The box A sequence from -30 to -40 is much less conserved and seems to be unique of promoters for stable RNA genes of Methanococcus. The single exception is the tetranucleotide AAAA, located at -35, which is also found upstream of stable RNA genes of Thermoproteus (Fig.4b). The box B motif ATGC is common to methanogens/halophiles and Sulfolobus but is not present in Thermoproteus. As with the consensus sequences of eukaryotic (5) and eubacterial (26) promoters, no position of the octanucleotide is invariant (Fig. 4c). The highest conservation shows the T in the third and the A in the fourth position, whereas A and T occur at position five almost with the same frequency. The subkingdom-specific variation of the consensus in the extreme halophiles (C at position one and G at position five instead of a T and an A; Fig. 4d) might reflect the high GC-content of these organisms (27). The distance of the first nucleotide in the consensus relative to

the RNA start site ranges from -25 to -32 (Fig. 4a and 4d). In 10 of 15 genes, however, it is located between -30 and -32, respectively (Fig. 4d). Thus this sequence is highly conserved at the same location in phylogenetically disparate organisms (8). On the basis of these data, we propose the overall consensus of the -25 region of archaebacterial promoters for stable RNA genes as: TTTA$^{A}_{T}$ATA (Fig. 4d).





c

T_{22}	T_{28}	T_{30}	A_{28}	A_{13}	A_{21}	T_{23}	A_{21}
C_6	A_3	A_1	T_3	T_{14}	G_7	A_4	G_5
A_2				C_3	T_2	C_4	C_3
G_1				G_1	C_1		T_2

d

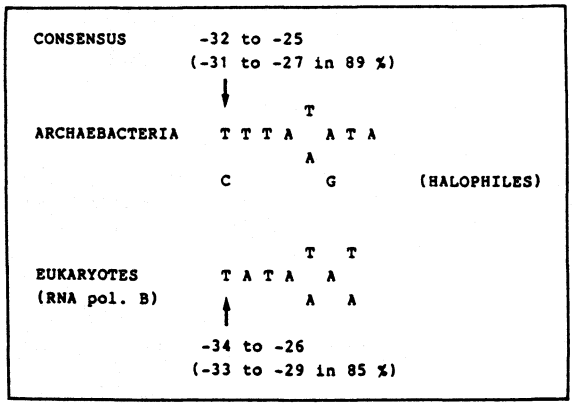


Fig. 4 RNA polymerase binding site and inferred archaebacterial consensus promoter sequence.
a) Limits of the RNA polymerase binding site. The nucleotide sequence of the 5' flanking region of an rRNA and tRNA gene, containing two sequences highly conserved between stable RNA genes of Methanococcus ((box A and box B; (12)), is shown. The boundaries of the promoters are marked by arrows.
b) Comparison of the upstream regions of archaebacterial genes. The sequence of the noncoding strand of 29 archaebacterial genes is shown. Alignment was performed with reference to the AT-rich -25 region within the RNA polymerase binding site (Fig. 4a). Nucleotides with homology to the derived consensus $TTTA^{T}_{A}ATA$ at -25 and ATGC around the RNA start sites (small letters) are underlined.
c) Consensus sequences in the region of transcription start sites of archaebacterial genes. The upstream sequences of the noncoding DNA strand of archaebacterial genes were aligned with reference to the conserved sequence $TTTATATA$. The subscripts denote the base frequency at each position.
d) Comparison of consensus promoter sequences. The AT-rich consensus characteristic of archaebacterial stable RNA and eukaryotic RNA polymerase B (5) promoter sites were aligned to yield maximum homology. The spacing of the first nucleotide in the conserved sequences relative to the known (18 genes; references 11-13, 23, 28, 29, 31) or putative (13 genes; references 12, 13, 22, 24) transcription start sites is indicated in each case.

Also the 5' flanking regions of the few sequenced archaebacterial protein-encoding genes (halophiles: bop (28), brp (29), cell surface glycoprotein (30); sulphur-dependent thermophiles: VP1 (31)) show at -25 considerable homology to the stable RNA consensus (Fig. 4b). This might indicate that this octanucleotide is also an element of promoters of protein-encoding genes.

Relation of archaebacterial, eukaryotic and eubacterial promoter sequences.

Promoters protected by the Methanococcus RNA polymerase contain an AT-rich sequence at position -25 and a second region of homology at the start of transcription. The absence of sequence homology to the eubacterial -35 sequence and the different spacing of the AT-rich region (Fig. 4d) relative to the start of transcription clearly discriminates between archaebacterial stable RNA and eubacterial promoters. Both the location at nucleotide position -25 and the sequence of the archaebacterial consensus octanucleotide are strongly reminiscent of the TATA box of promoters preceding highly expressed protein-encoding genes of higher eukaryotes (5) (Fig. 4d), whereas the location differs from that of TATA elements in yeast promoters which are arranged between 40 and 120 bp upstream of mRNA initiation sites (32). The homology of the -25 region in archaebacterial and eukaryotic promoters would be in line with a specific phylogenetic relationship between archaebacterial RNA polymerases and RNA polymerase B. Direct evidence for this relationship has been indeed provided by sequence analyses of RNA polymerase genes (33).

ACKNOWLEDGEMENTS

We thank Drs. August Böck and Karl O. Stetter for valuable discussions and Drs. Gary Sawers, Manfred Sumper and David Ubben for comments on the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft to A. Böck, K.O. Stetter and M. Thomm.

REFERENCES

1. Woese, C.R. and Fox, G.E. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5088-5090.
2. Schaller, H., Gray, C. and Herrmann, K. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 737-741.
3. Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 784-788.
4. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
5. Corden, J., Wasylyk, B., Buchwalder, A, Sassone-Corsi, P., Kedinger, C. and Chambon, P.. (1980) Science 209, 1406-1414.
6. Hall, B.D., Clarkson, S.G. and Tocchini-Valentini, G. (1982) Cell 29, 3-5.

7. Verbeet, M.Ph., Klootwijk, J., van Heerikhuizen, H., Fontijn, R.D., Vreugdenhil, E., and Plantu, R.J. (1984) Nucl. Acids Res. 12, 1137-1149.
8. Woese, C.R. and Olsen, G.J. (1986) System. Appl. Microbiol. 7, 161-177.
9. Huet, J., Schnabel, R., Sentenac, A. and Zillig, W. (1983) EMBO J. 2, 1291-1294.
10. Hamilton, P.T. and Reeve, J. (1985) Mol. Gen. Genet. 200, 47-59.
11. Chant, J. and Dennis, P. (1986) EMBO J. 5, 1091-1097.
12. Wich, G., Hummel, H., Jarsch, M., Bär, U. and Böck, A. (1986) Nucl. Acids Res. 14, 2459-2479.
13. Wich, G., Leinfelder, W. and Böck, A. (1987) EMBO J. 6, 523-528.
14. Dennis, P. (1986) J. Bacteriol. 168, 471-478.
15. Thomm, M. and Stetter, K.O. (1985) Eur. J. Biochem. 149, 345-351.
16. Brosius, J. (1984) Gene 27, 161-172.
17. Shalloway, D., Kleinberger, T. and Livingston, D.M. (1980) Cell 20, 411-422.
18. Linxweiler, W. and Hörz, W. (1982) Nucl. Acids Res. 10, 4845-4859.
19. Brosius, J., Dull, T.J. and Sleeter, D.D. (1981) J. Mol. Biol. 148, 107-127.
20. Buttner, M.J. and Brown, N.L. (1985) J. Mol. Biol. 185, 177-188.
21. Lewis, M.K. and Burgess, R.R. (1980) J. Biol.Chem. 255, 4928-4936.
22. Daniels, L.J., Douglas, S.E. and Doolittle, W.F. (1986) System. Appl. Microbiol. 7, 26-29.
23. Moritz, A., Lankat-Buttgereit, B., Gross, H.J. and Goebel, W. (1985) Nucl. Acids Res. 13, 31-43.
24. Larsen, N., Leffers, H., Kjems, J. and Garrett, R.A. (1986) System. Appl. Microbiol. 7, 49-57.
25. Reiter, W.D., Palm, P., Voos, W., Kaniecki, J., Grampp, B., Schulz, W., Zillig, W. (1987) Nucl. Acids Res. 15, 5581-5595
26. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
27. Jones, W.D., Nagle, D.P.Jr. and Whitman, W. (1987) Microbiology Reviews 51, 135-177.
28. DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 125-129.
29. Betlach, M., Friedman, J., Boyer, H. and Pfeifer, F. (1984) Nucl. Acids Res. 12, 7949-7959.
30. Lechner, H. and Sumper, M. J. Biol. Chem., in press.
31. Reiter, W.D., Palm, P., Yeats, S. and Zillig, W. (1987) Mol. Gen. Genet. 209, 270-275.
32. Struhl, K. (1987) Cell 49, 295-297.
33. Pühler, G., Leffers, H., Garrett, R. and Zillig, W. pers. communication.