Nucleic Acids Research, Vol. 20, No. 20 5423–5428

Elements of an archaeal promoter defined by mutational analysis

Johannes Hain*, Wolf-Dieter Reiter+, Uwe Hüdepohl and Wolfram Zillig
Max-Planck-Institut für Biochemie, 8033 Martinsried, Germany

Received June 24, 1992; Revised and Accepted September 25, 1992

ABSTRACT

The sequence requirements for specific and efficient transcription from the 16S/23S rRNA promoter of *Sulfolobus shibatae* were analysed by point mutations and by cassette mutations using an *in vitro* transcription system. The examination of the box A-containing distal promoter element (DPE) showed the great importance of the TA sequence in the center of box A for transcription efficiency and the influence of the sequence upstream of box A on determining the distance between the DPE and the start site. In most positions of box A, replacement of the wild type bases by adenines or thymines are less detrimental than replacements by cytosines or guanines. The effectiveness of the proximal promoter element (PPE) was not merely determined by its high A+T content but appeared to be directly related to its nucleotide sequence. At the start site a pyrimidine/purine (py/pu) sequence was necessary for unambiguous initiation as shown by analysis of mutants where the wild type start base was replaced. The sequence of box A optimal for promoter function in vitro is identical to the consensus of 84 mapped archaeal promoter sequences.

INTRODUCTION

Two consensus regions have been defined by comparison of promoter sequences of Archaea [1] (Archaebacteria): the box A centered about 27 bases upstream of the transcription start site and the box B at the start site [2, 3].

Mutational analysis of the 16S/23S rRNA promoter of *Sulfolobus shibatae* using an *in vitro* transcription system [4] has identified the position and function of three essential promoter elements: (i) a distal promoter element (DPE) encompassing box A, which is important for transcription efficiency and start site selection; (ii) a proximal promoter element (PPE) between position −11 and −2 which also contributes to transcription efficiency; and (iii) a pyrimidine/purine-sequence which, at the proper distance from the DPE, serves as start site [5]. Utilisation of heterologous promoters in the *S. shibatae* transcription system showed that constitutive promoters are functionally conserved between distantly related archaea [6]. An analysis of the *Methanococcus vannielii* $tRNA^{Val}$ gene promoter [7] is in fair agreement with these data but extends the determination of sequence requirements by point mutations in which certain bases between position −35 to +2 were replaced with guanine.

In this study we analyze the functional importance of positions in promoter elements in more detail, especially within box A where each base was replaced with the other three. We also examined the effect of all possible base exchanges at the start site in order to check the requirement for a pyrimidine/purine sequence. Moreover, we replaced the proximal promoter element by streches of adenines or thymines, or the complement of the wild type PPE-sequence, to test the hypothesis that a high A+T content is sufficient for its function in the 16S/23S rRNA promoter [5]. Promoter strength and transcript start sites of all mutant promoters were determined by S1 nuclease mapping of the *in vitro* transcripts. The box A sequence optimal for promoter function in vitro was compared to the consensus of mapped archaeal promoters.

MATERIALS AND METHODS

Materials

Restriction enzymes, RNase-free DNase and T4 ligase were obtained from Boehringer Mannheim, T4 polynucleotide kinase, S1-nuclease and Klenow-fragment of E.coli DNA polymerase I from Pharmacia, radiochemicals from Amersham. The soluble cell-free extract of *S. shibatae* was prepared as described [4].

Methods

Construction of the vector and mutants. A sequence identical to the −39 to +6 promoter region of the *S. shibatae* 16S/23S rRNA operon [14] *was cloned into the phagemid pBluescript II KS+* (Stratagene) between the *Xho* I and *Bam* HI cleavage sites, using synthetic oligonucleotides. The *Pst* I-site in the middle of the promoter region was generated by exchange of a thymidine against guanosine at position −19 and against cytidine at position −21 in the oligonucleotide sequence (Fig. 1). This difference in the promoter sequence had no negative effect on transcription efficiency and start site selection. This vector construct was named pSP2 and used as standard. The *Pst* I-site together with either the *Xho* I- or the *Bam* HI-site were used to generate mutants by introduction of synthetic oligonucleotides comprising distinct mutations. These oligonucleotides were designed with ends

* To whom correspondence should be addressed

+ Present address: MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824-1312, USA

complementary to the sticky ends of either the *Xho* I/*Pst* I-or *Pst* I/*Bam* HI-vector fragment generated by digestion of pSP2 with the respective enzymes. Both double digests always led to a vector fragment with overhanging ends on the same strand (Fig. 1). Ligation [8] of the oligonucleotides with these vector fragments led to a circular vector with a short single stranded region. Upon transformation [9] the DNA-repair system of the recipient-cell strain E.coli XL1-blue converted the construct into double-stranded DNA.

After selection on LB agar containing tetracyclin (25 μg/ml) and ampicillin (75 μg/ml) the recombinant DNA was isolated [10] and the mutation verified by DNA-sequencing [11]. Mutated phagemids obtained through this procedure were cleaved with *Bgl* I yielding two fragments and used as linear templates for in vitro transcription.

In vitro transcription. 100 ng template-DNA were 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 CTP, 1 mM GTP, 1 mM UTP and 8 μl of the cell-free extract from *S. shibatae* [4]. The mixture was chilled on ice, 50 μl ddH_2O were added and the reaction mixture was extracted three times with 100 μl of phenol/chloroform/isoamylalcohol 25:24:1 (vol/vol/vol). The nucleic acids were precipitated with ethanol and the template-DNA was removed by digesting with 25 U RNase-free DNase I (Boehringer, Mannheim) in 50 μl reaction buffer for 30 min at room temperature. After addition of 50 μl ddH_2O the reaction mixture was extracted twice with phenol/chloroform/isoamylalcohol and with chloroform/isoamylalcohol 24:1 (vol/vol). The aqueous phase containing the *in vitro* RNA was stored at −70°C.

S1 analysis. DNA probes for S1 nuclease analysis [12, 13] were prepared by extension of the 5'-^{32}P-labeled M13 universal sequencing primer hybridized to single stranded pSP2 or its respective mutant derivative single strand. Identical volumes (two microliters) of the *in vitro* RNA solutions were hybridized to an at least fivefold molar excess of DNA probe. Hence only the quantity of the *in vitro* RNA determined the strength of the signal and therefore the transcription efficiency. The condition for hybridization, S1-nuclease digestion and the electrophoresis through denaturing polyacrylamide gels were as described [4].

The transcription efficiencies were quantified by densitometry of the autoradiographs. A calibration curve was used to correct for nonlinearity.

Primer extension. After *in vitro* transcription and DNase I digestion the reaction mixtures were extracted twice with phenol/chloroform/isoamylalcohol and were ethanol coprecipitated with 30,000 cpm of the 5'-^{32}P-labeled M13 universal sequencing primer. Reverse transcription was performed as described [14] yielding a fragment of 99 bp (in the case of initiation at the wild type start site). The cDNA was analyzed on a 6% polyacrylamide sequencing gel.

RESULTS

In order to define the sequence requirements within the archaeal promoter elements responsible for transcription efficiency and start site selection, an extensive mutational analysis of the promoter region of the 16S/23S rRNA operon of *S. shibatae* was performed. For this purpose a cassette with a sequence almost

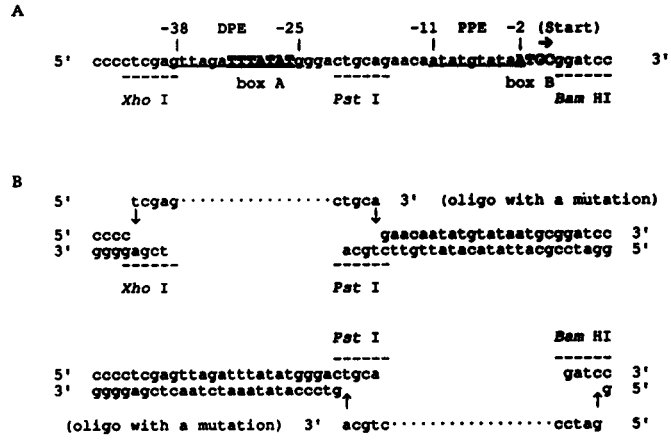

Figure 1: (A). Sequence of the promoter region of the 16S/23S rRNA operon of *S. shibatae* containing a *Pst* I site introduced by two nucleotide exchanges at positions −19 (T to G) and −21 (T to C). Oligonucleotides with this promoter sequence from position −39 to +6 (transcription start site position +1) have been inserted into pBluescript II KS+ between the *Xho* I and the *Bam* HI sites yielding the construct pSP2. The sequence shows the distal promoter element (DPE, underlined) including box A (uppercase and bold), the proximal promoter element (PPE, underlined) and box B (uppercase and bold). The M13−20 universal primer used for sequencing, generation of the S1 probes and for primer extension binds at position +83 to +99. **(B).** Strategy of the insertion of synthetic oligonucleotides containing certain mutations into pSP2. The pSP2 construct was cleaved with either *Xho* I/*Pst* I (upper part) or with *Pst* I/*Bam* HI (lower part). The synthetic oligonucleotides were complementary at both ends to these restriction sites. After ligation and transformation the single stranded region was filled in by the DNA repair system of the host cell.

identical to that in the natural promoter from position −39 to +6 (transcription start site defined as +1) was assembled from overlapping oligonucleotides and cloned into pBluescript II KS+ yielding pSP2 (Fig. 1). To permit the facile and efficient manipulation of all parts of the promoter, the pSP2 promoter cassette contained a *Pst* I site between the distal and the proximal promoter element leading to two nucleotide exchanges compared to the wild type promoter. *In vitro* transcription of pSP2 and a similar construct containing an entirely wild type promoter sequence indicated that the introduction of the *Pst* I site did not alter the initiation site or negatively affect transcription efficiency (data not shown). Mutations were focussed on three regions: (i) box A (position −32 to −26), where each of the seven bases of the wild type sequence was substituted by each of the other three possibilities (figure 2, upper part), (ii) the PPE, which was examined by three cassette mutants (figure 2, middle), and (iii) box B with the start site, where the wild type guanine at position +1 was substituted by each of the three other bases (figure 2, bottom).

The mutations in the promoter region were introduced by ligation of synthetic oligonucleotides and verified by DNA sequence determination. The various constructs acted as template for *in vitro* transcription with a cell free extract of *S. shibatae*. The transcription efficiency of each mutant promoter was determined by S1-nuclease mapping of the resulting transcription products followed by densitometric analysis of the respective signals on the autoradiogram. To check the reproducibility for each individual construct, at least three independent repetitions of these experiments were performed.

In the case of the box A mutants, a pSP2-derived probe was

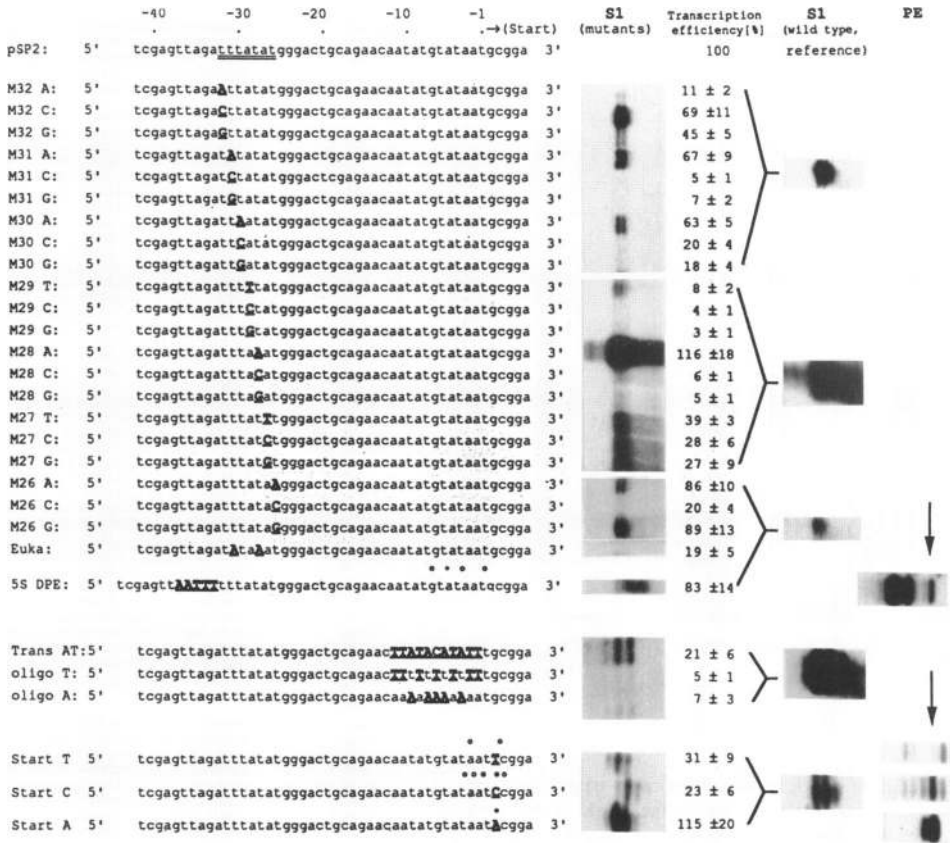

Figure 2: Transcription efficiencies of all mutant constructs mapped by S1-analysis. The data were derived from at least three independent repetitions of the S1-analysis. Box A-mutations upper part, PPE-mutations in the middle and box B-mutations at the bottom of the list. Mutations are shown by bold underlined uppercase letters, wild type nucleotides in lowercase letters. The box A motif defined by functional analysis is double underlined. The autoradiographs from S1 experiments and the deduced transcription efficiencies (± standard deviation) are shown to the right of the template sequences. Autoradiographs shown were assembled from several gels; the corresponding wild type (=pSP2) controls are shown to the right of the transcription efficiency data. Autoradiographs from the start site determinations by primer extension (PE) are shown to the extreme right of the figure; arrows indicate the fragment corresponding initiation at the wild type +1 position. Mapped start sites are also indicated on the template sequences, if deviating from the wild type start site position: major initiation site: black dot; minor initiation site(s): open circle.

used in S1 analysis for the examination of the transcription efficiency and the start site. Since the probe was complementary to the template DNAs up to box A, the start sites of all these mutants could be monitored without difficulty. In case of mutations in the PPE and in box B promoter mutants, specific DNA probes were prepared for each construct to exclude S1 nuclease digestion at mismatch positions.

DPE mutants

In the following part, the influence of mutations in box A of the 16S/23S rRNA promoter shall be described in the order of decreasing effects on transcription efficiency. All nucleotide exchanges at positions −30 and −29 reduced transcription efficiency dramatically (Fig. 2), indicating the importance of the TA-sequence at these positions for the function of box A; only the adenine at position −30 maintained a high transcription efficiency. Every exchange at position −27 also led to a reduction, though to a lesser extent than at the positions −30 and −29. The thymine at position −32 could only be replaced by cytosine, the thymines at positions −31 and −28 only by adenines, in each case leaving the transcription efficiency higher than 67% of that of the wild type promoter (Fig. 2). Guanine at position −32 was tolerated, but in a lesser extent (45% transcription efficiency). The other possible exchanges at each of these positions left, at most, 11% of the transcription efficiency. The tolerance of the transcription system especially towards cytosine, but not adenine or guanine, at position −32 shows that at this position occupation by a pyrimidine rather than a high A+T content is necessary for promoter function. In contrast, the tolerance towards transversion at positions −31 and −28 suggests that a weak base pairing rather than a certain base is required there. At position −26, only the replacement of the wild type thymine by cytosine led to a strong reduction in transcription efficiency: the other two possible exchanges were tolerated (Fig. 2; see also figure 3 for an overview). All of these box A mutations influenced the transcription efficiency without leading to a shift of the transcription start site.

One mutant ('5S DPE', Fig. 2) contained the DPE of the 5S rRNA gene promoter of *S. shibatae*. This alteration in the region immediately upstream of the box A, which itself was the same as in the 16S/23S rRNA promoter, left the transcription efficiency at 83% but introduced ambiguity of transcription initiation. Starts were mapped at positions −7 (guanine), −5, −3 (adenines) and +1 (guanine) with the major start site at position −5.

The mutation 'Euka' transformed the wild type box A sequence 'TTTATAT' to 'TATAAAT' which represents the 'TATA-box' consensus of eukaryotic pol2 promoters. The transcription efficiency from this promoter mutant was 19% and thus

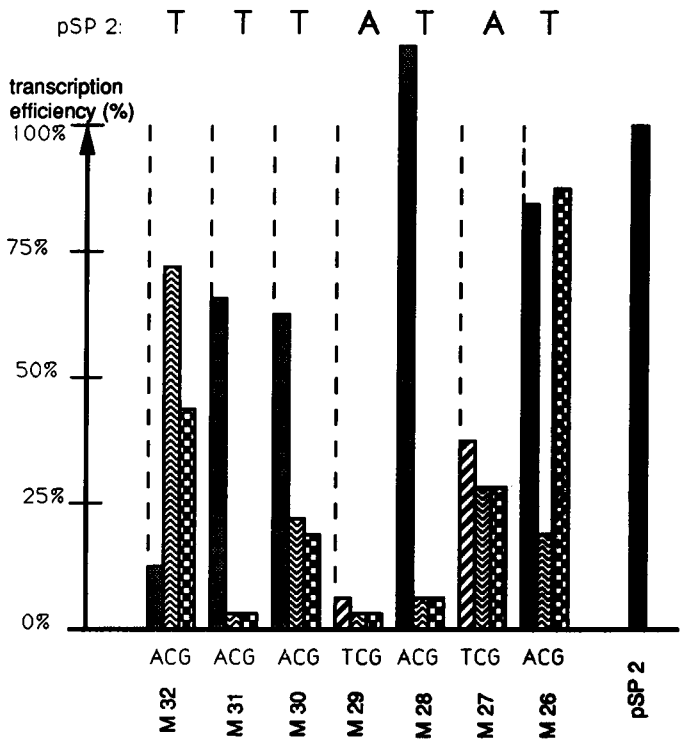

Figure 3: Transcription efficiencies of box A mutants obtained by quantitative S1-mapping. The names of the mutants are given below the x-axis at their corresponding column and are in accordance with those at figure 2. The black column at the right represents the pSP2 (the 16S/23S rRNA operon promoter introduced into pBluescript II KS+).

significant though weak. Transcription was initiated at the same site as in the wild type promoter.

PPE mutants

The three cassette mutations in the PPE, an oligo adenine stretch, an oligo thymine stretch and the complement of the wild type PPE-sequence, reduced the transcription efficiency to 7%, 5% and 21%, respectively. The start site of each of these three mutants was the same as that of the wild type promoter (Fig. 2).

Box B mutants

In box B the guanine at the transcription start site was substituted by each of the three other bases. The replacement of the wild type guanine by adenine did not shift the transcription start site but increased the transcription efficiency slightly. The replacement by thymine shifted transcription initiation to the positions −3 (adenine) and +2 (cytidine) accompanied by a reduction of transcription efficiency to 31% overall. The replacement by cytidine led to an ambiguous initiation in the vicinity of the wild type start site but mainly at the position +1 (cytidine) with an overall transcription efficiency of 23% (Fig. 2).

DISCUSSION

Using an *in vitro* transcription system from *Sulfolobus shibatae* [4], the functional significance of single positions within archaeal promoter elements was examined by determining the strength of altered promoters and their start site selection in comparison to the wild type promoter of the 16S/23S rRNA-operon of *S. shibatae*. The latter was chosen because of its strength *in vivo*

organism and gene	sequence	reference
H.c. rRNA P1	TCGACGGTGTTTTAAGTACCCCA CCACTCGGAT GAGATGCGAA CGA	[16]
H.c. rRNA P2	GTCCGATGCCTTTAAGTACAACA GGGTACTTCG GTGGAATGCG AAC	[16]
H.c. rRNA P3	ATTCGATGCCTTTAAGTAATAAC GGGTGTTCCG ATGGAATGCG AAC	[16]
H.c. L11e	AAAGACAAGGGTTAAACCCGCGG CGGCGGTTTC TCGGNGTGTG GCT	[17]
H.c. L1e	CTTCGACGCTTTTAAGCCCGGGA TCACCGTCTG TAGAACCGGA ACA	[17]
H.c. NAB	GTTTCGACACGTTAAGACGCCGA GTGAAGCCAT CGCATAGTGA TGG	[17]
H.c. SOD	CGGAAACCACCATAAGCAGCGCC GACGTACGNC ACACTGTATG TCC	[18]
H.h. RNAP	GTTGACAAGGCTTAAATGCTGTG GGCAGCAACT GGCCTGTGGT ATC	[19]
H.h. S12	AAAGGTCGGGCTTAAGTGCCTCC GCGGGATACT CGGCTGTGTG ACG	[19]
H.h. HOP	GTTGGGGGAGGTTATTTAATGGC GTGCCGTGTC CTTCGAACA CAT	[20]
H.h. BOP	GGGTCGTAGAGTTACACACATAT CCTCGTCAGG TACTGTTGCA TGT	[21]
H.h. Glycopr	GCAGAAAGCATTTACCAGTGGCC GGGTATAGTC TGGAGCGCCC CTA	[22]
H.h. flgA	GGGACAAACGTTTATTAGTCAGC TGTCACACTC ACGTCAAACA CTC	[23]
H.h. flgB	CCTACACTTTGGTATCGATGGCC GATCTGTATG GGTGAGCCCC AGC	[23]
H.h. Mn-SOD	CTCACGAACATTTAACATGACGC CGCGTGATCA CTGATCCGGT GGA	[24]
H.h. p-vac	CATACACATCTTTTAGTGATGCC CGAGTATAGT TAGAGATGGG TTA	[25]
H.h. p-gvpD	GAGTCGTACTTCTAAGTACGGAG AGTGTAAAGC TTCTTAGCA GTC	[26]
H.h. ΦH T1,2,3	GAGGGTCAATTTTATATACTGG GGTTCCAGGG ACATGACGGA GCA	[27]
H.h. ΦH T4	TAGGAATAGATATAAGTTAGACC CCTCGTAAAG TCCAGCTGA CGA	[27]
H.h. ΦH T6	ATGGGAACACGTTATGATGGCC AAAAACCTCT TTTAGGTCGT GCA	[27]
H.h. ΦH T7	CTGACGGGGATTTATCTTTGCAC GCATGGAAGT CCACTCGTTC CAT	[27]
H.h. ΦH T8	ATCAAACAGATTTAATAGGTAGG GGCCTCTAAT GTTTGACGAG GTA	[28]
H.h. ΦH T9/10	TGGCACGAAATTTACTTACCGAT GCGTCTATTC TATTCTGCA TGA	[29]
H.h. ΦH T1SH1.8	CGTCACAAGAGTTATCTGAATTG GGTGTCTGGT ATCTGCTGAG GCC	[30]
H.h. ΦH TLH1	CGGAAAATCGTTTTAGCGGATGCA GTGTACTCCG AGGCAGCAC GAG	[31]
H.h. ΦH T_{ant}	CGTAACCTTTGTTACCTTGCGAA TCGTGGTTTG GCATGCTCTG GGC	[29]
H.mm. 16S/23S rRNA P1	CTTCGACGGCGTTAAGTGTGGCT CACCCATGGG AATGAAATGC GAA	[32]
H.mm. 16S/23S rRNA P2	TTCCGACGCCCTTAAGTGTAACA GGGCGTTCGG AATGAACGCA AAG	[32]
H.mm. 16S/23S rRNA P3	ATCCGACGCCCTTAAGTGTAACA GGGTGCTCGA ATGAACGCGA ACG	[32]
M.t. mvhDGAB	ATAAGTAACCTTTATGCTTACCA CTCCTAAACC ATAATGCGAG TAC	[33]
M.t. mcr	AGAAAGAAATCTTAATTAATTAT GATCAAGTTA AATAGCAGGC GTA	[34]
M.t. purE	GGTGCCATCTTATAAATATGACC TCCGATAGGT TCTCTGTATA TAA	[35]
M.v. 16S/23S rRNA	TACCTAAAACAATACATATTACA ACACGTTTTC ATATTATGCA AAT	[15]
M.v. tRNA & 5S rRNA	TACCGAAAACTTATATATTATA ACACTAGTAT TCAGTATGCG AAC	[15]
M.v. $tRNA^{Arg}$	AACCGAAATATTTATATACTAGA ATACCCTTCC TATACTATGC TCT	[15]
M.v. ORFa	TAAGCATAATTTTTAAATACCATT AACCAGTATG TCTATCGTAT AGA	[36]
M.v. S17	CTATATTAAACTTATATACCTGT AGTTCGTATT AAATCTGGAA CTT	[36]
M.v. hisA	TAGGTACCAATATATGTTTAAA ACCTAATTTA ACATAGTTTT AAT	[37]
M.v. mcr	TGAAAACTTGAATATATCTTTCCT TTAATAATGT TATGATGTAT ATA	[38]
M.v. ORF1	TATGCAAAAGTATAAAAATAACA TCTAACTATT AACTTACTCG TAT	[39]
T.a. 16S rRNA	GCCTTCGAAAGTATATATACTG ATTTGCTATT CTTTACTTTG CAC	[40]
T.a. 23S rRNA	GATCAAAATGCTTATATCCCTCT TAATGATATA GTCCATGCAC GCT	[40]
T.a. 5S rRNA	TCACGAAAATCTTATATAGATGT GTTCTATATA GTGTCGGCA ACG	[40]
T.a. $tRNA^{Met}$	CCCTACAGCCATTATTCGTCCAC TATGCTGGAT TGAAGCGGGG TGG	[41]
D.a. Ligase	CATCATCAAATATATTATATCGT AATTCAAACT TTTATTTGTG GAG	[42]
D.a. SOR (aer)	AAAGAAAGAAATATAAAAGTAGNC AGAAATGTAT ATTTGACCGA AAA	[43]
D.a. sor-ORF2 (konst.)	AATTCACTTATTTATACATTTGT TAAGGATGGG TTATATTATG CAT	[43]
D.a. sor-ORF3 (aer)	GAAGAATAGATTTATTTCCTGAA TAAAGTAGAT TGATATATGC TCG	[43]
D.a. sor-ORF4 (anaer)	CTCAACTATTTTTAAATACTAGT GCAAGAAAGA ATAGCATATG AGT	[43]
D.m. 5S rRNA	CCTAACACACTATACAATATATT GATGCTCGCA ATAGTGGTGA CCC	[44]
D.m. rRNA-P1	TATATAAACCCTTATCGCATAGA GTAAGATTCC AGACGCTTAC AGC	[44]
D.m. rRNA-P2	ACCCGTCATGATTAATACCCTTG GAGCAAATAG ATTCATCTAA CGG	[44]
S.a. S12	CTAAGTTAATATTAAAAAGGTAT TTGGATATTA ATTCTGTTGT TGT	[45]
S.a. ORF88 (rpoH)	CTAACATAATTTTATTAAACTTA AATGAGTAGG ATGAGATTGC TGT	[45]
S.a. rpoC	CTGTGAGAAGTTTATATGGAGAT ATTGTTCAAG TAGTAGTGGG TGA	[45]
S.a. ORF-X	CGGTAAACCTTTTATGGTATCTG TTATTAGGTA TAGTTGATGA AGG	[45]
S.a. SOD	AATCAAAATATTTAATCCACCAT TGGCTATTGG TTTGTGGAAC CCA	[46]
S.a. S7	CACAAACTCTATTAAACCCTTCA TAATATATTT AAGGTGGCAG TAA	[47]
S.a. $tRNA^{Ser}$	TTTGATTATTTTTATAAGCATCT CGTATAATCT TATTAGTGGC CGG	[47]
S.a. Ef 1a	TAGCTAAACTTTTAATGCTAGTC TAGTTATAAA AATGTGAAAG GTA	[48]
S.a. Ef 2	TCCAACACACTTTAATACTTAAA AGTTAACTGA CTATTGGAA CTG	[48]
S.s. 16S/23S rRNA	AGAAGTTAGATTTATATGGGATT TCAGAACAAT ATGTATAATG CGG	[2]
S.s. 5S rRNA	TAGTTAATTTTTTATATGTGTTA TGAGTACTTA ATTTTGCCCA CCC	[2]
S.s. $tRNA^{Arg}$	TGTCTAACAGTTTAAAATATCCG ATGAGGCATA TTTATGTTGC ACC	[49]
S.s. SSV1 T1,2	ACTGGAGGGGTTTAAAAACGTAA GCGGGAAGCC GATATTGACC AAG	[2]
S.s. SSV1 T3	TTAGGCTCTTTTTAAAGTCTACC TTCTTTTTCG CTTACAATGA GGA	[2]
S.s. SSV1 T4	GATAGCCCTTTTTTAAAGCCATAA ATTTTTTATC GCTTGATGAA GTG	[2]
S.s. SSV1 T9	AAGTAGGCCCTTTATAAAGTCAT ATTCTTTTTC TTTCCCTGAT GAG	[2]
S.s. SSV1 T5	TAGAGTAAGACTTAAATACTAAT TTATACATAG AGTATAGATA GAG	[2]
S.s. SSV1 T6	TAGAGTAAAGTTTAAATACTTAT ATAGATAGAG TATAGATAGA GGG	[2]
T.c. 5S rRNA	GACCGTAACCTTTATAAACCCAG GCCGGGAGTT CCGTACCGGT ACG	[50]
T.c. rpoH(1)	TCGTAACCGATATAAACCCCTCT GAGCGAGTTG TCACTCGAAA AGC	[51]
T.c. rpoH(2)	ACTCGAAAAGGTTATATACCGAA ACCCGCAGGG GGAGGTAGTG ATA	[51]
T.c. rpl30	ACCCGAAAAGGTTAAAAAGCACC GCCTTAAGGG TCTTTCGCGC GTC	[52]
T.c. rps12	CTAGGAACATTTTATAAGAGCCT CGGGGTTAAG GTTAAAGGGG TCC	[52]
T.c. $tRNA^{Thr}$	GAGGGAAAAATTTATAAATGCCA ATCTGCATGC TATCATCGAC GCC	[46]
T.c. $tRNA^{Pro}$	GAGGGAAAAGCTTAATAAGGAAC CCTTTAACCC ACTACCGCGG CATG	[46]
T.p. 16S/23S rRNA	GCATAATATTCATATAACCCCCC GTTTACTAAC TAGATTGCCG CCA	[53]
T.p tRNA^{Met}	ATGCTAAAGGTTTATTACCCAGG AAGTATTCCG GTCATGGGGG GTT	[53]
T.p. ORF1	GCATTCGACAAATATTTTAACCT CGTTCGAGGA GTATTGCAGGT GGT	[53]
T.p. ORF2	TAAATAGCTTTTTTAACGTAGGAG GTGCTACGTC CACGGGATCG TAA	[53]
T.t. 16S/23S rRNA	GCGAAAAATTTTTAATTAGGGT GTTTTAGGAT GGTCGCGCCT TAA	[54]
T.t. $tRNA^{Ala}$	AGCGAAAAAATTTAAATCGGTGA GTAAGTACGC TCGGGCCCGG TAG	[54]
T.t. $tRNA^{Met}$	ACAAAGCTTTTTTAAATTCGCGC AAAGCTTAGA CCTGGCGGGG TAG	[54]
consensus:	$^{T}_{C}TTA^{T}_{A}A$	
H.c. ORF	GGAAAGCGCTTTTCGGCGCTTGCTGTCTACGGGCACGTGATG	[17]
H.h. BRP	GGTCTTTTTTGATGCTCGGTAGTGACGTGTGTATTCATGTGACCA	[55]
H.h. ΦH TLX3	CGATGGGTTCAAGAACCTCCCTGCCCGTCTATTTCAATGGATGCCG	[31]
H.me. mc-gvpA	CACGAATGATTTTGTTACTTGCCAACACGTTTTCAGATGGGTA	[56]
S.s. SSV1 T_{ind}	GTCGACTCTGTGTATCTTATGTATCTTATACAAAAAATATGGGA	[2]

Figure 4: The list (first part euryarchaeotal, second part crenarchaeotal species) shows the transcription start sites of mapped archaeal genes and their promoter sequences (box A-sequences are underlined and the transcription start sites bold and underlined). The sequences are aligned for the box A element, except the last five sequences which are aligned for the transcription start sites since no consensus box A could be found. Abbreviations used for organisms: H.c.: Halobacterium cutirubrum H.h.: Halobacterium halobium,, H.mm.: Halobacterium marismortui, H.me.: Halobacterium mediterranei, M.t.: Methanobacterium thermoautotrophicum, M.v.: Methanococcus vannielii, T.a.: Thermoplasma acidophilum, D.a.: Desulfurolobus ambivalens, D.m.: Desulfurococcus mobilis,S.a.: Sulfolobus acidocaldarius, S.s.: Sulfolobus shibatae, T.c.: Thermococcus celer, T.p.: Thermofilum pendens, T.t.: Thermoproteus tenax.

and *in vitro* and its identity to the promoter consensus defined by sequence comparison [2, 3, 15].

Promoter efficiency

The box A. All box A mutations replacing the thymine at position −30 and the adenine at position −29 in the center of box A showed strong reductions of transcription activity indicating the importance of the TA-sequence at these positions. The only exception, a tolerance towards the exchange of thymine −30 against adenine, can be interpreted as a shift of the TA-sequence one position upstream. These data are in accordance with a high conservation of the TA-sequence in the center of box A in archaeal promoters (Fig. 4). The distance between the TA-sequence and the start site is 29 bases in the 16S/23S promoter but between 25 to 28 bases on average.

These results, and the sensitivity of the promoter to certain base exchanges at single positions of box A, defined an optimal box A sequence for promoter function. For the determination of this optimal box A sequence, only mutants with a transcription efficiency higher than 66% of the wild type were considered. Position −26 of the 16S/23S promoter was not considered important for promoter function since only one of the three possible base exchanges, the introduction of cytosine instead of the wild type thymine, led to a strong reduction of promoter strength. Applying the above criteria, an optimal box A sequence 5′ T/CTTAT/AA 3′ (positions −32 to −27) was derived. This functionally-determined sequence was in good agreement with the consensus of archaeal box A sequences 5′ TTTA$^{T}/_{A}$A 3′ [2, 3, 15]. Inspection of 84 mapped archaeal promoters (Fig. 4) showed that cytosine did indeed sometimes replace thymine in the first position of the consensus box A. The last position, −27, in the consensus occupied by adenine proved very sensitive against base exchange in accordance with its conservation in all but the halobacterial promoters.

The PPE. Box A is not the only element determining the efficiency of transcription. Exchange of a second essential promoter element, the A+T rich PPE at positions −11 to −2, by stretches of adenines or thymines (Fig. 2) led to a 20 fold reduction in promoter efficiency and replacement of this element by its complementary sequence (Fig. 2) resulted in a 5 fold reduction of promoter strength. Thus, the element is not merely an A+T-rich region facilitating strand separation during the initiation of transcription, as previously discussed [5]. A certain sequence, the alternating purine/pyrimidine sequence 5′ ATATGTATA 3′ in the case of the strong 16S/23S rRNA promoter of *S. shibatae,* appears to be required for maximal promoter strength. The strong reduction of transcription efficiency upon replacement of the element by its complementary sequence indicates that the sequence must be correctly positioned with respect to the start site. The PPE sequence is not generally conserved between archaeal promoters (Fig. 4) and may therefore represent a particular feature of the 16S/23S rRNA promoter.

The start site. Promoter strength appears to depend on the distance between the DPE and the start site [5] as well as on the occupation of this site. Apart from the elements already discussed, the core promoter region thus includes the start site itself. When an unfavourable start site was introduced, e. g. a pyrimidine instead of the wild type purine, the start shifted to nearby sites at less favourable distances concomitant with a reduction of total transcription efficiency. In these cases, it is difficult to estimate the contributions of the changes of distance and start site context to the overall effect. But it appears that promoter strength declines sharply whenever a dinucleotide different from a pyrimidine/purine (py/pu) occupies the optimal initiation region defined by its distance from the DPE.

Start site selection

Analysis of promoter mutants with regard to start site selection indicated the necessity of a start motif as well as a distance measurement in defining a start region in which this motif serves its role. We suggested previously [5] that a purine preceeded by a pyrimidine acted as minimal start signal since most initiations on a number of mutant constructs occured at a purine following a pyrimidine. Results from a methanogene transcription system [7] and those of our current work corroborate this assumption. Furthermore a comparision of mapped archaeal transcription start sites showed that 79 of 89 transcripts initiate at a purine after a pyrimidine (Fig. 4). Since such a dinucleotide sometimes occurs more than once at an appropriate distance from box A and initiation nevertheless remained specific there must be additional information to provide specific initiation. The formerly proposed box B consensus $^{T}/_{A}TG^{C}/_{A}$ found around the start site [2] does not appear to play this role since initiation occurs either upstream, or within, or downstream, and thus not at a defined site in this sequence. Moreover many of the mapped promoters in figure 4 do not show this consensus. Aligning promoter sequences for the start site yielded a different, rather weak consensus $^{A}/_{T}T/_{C}G/_{A}$ with the initiation at the last position and in crenarchaeotal promoters a weakly conserved $(py/pu)_4$ pattern (see above).

The DPE as well as the py/pu dinucleotide determines the start site. This had already been shown with the insertion and deletion mutants [5] and was reconfirmed by some of our promoter mutants. Alteration of the distance between the DPE and the wild type start site led to a shift of the start site and concomitantly sometimes to ambiguity of initiation. This ambiguity was also observed when the wild type start guanine was substituted with thymine or cytosine. The mutant containing the 5S rRNA promoter DPE instead of the 16S/23S rRNA promoter DPE showed both an upstream shift of the start site and ambiguity in start site selection. The three different types of mutants share one common feature: all of the new start sites were positioned within a region of eight bases. This indicates that the DPE delimits (by some sort of distance measurement) a certain window of about eight bases in which initiation occurs.

Furthermore, the mutant with the 5S DPE shows that a stretch of more than four thymines upstream of box A altered the structure of the promoter in such a way that the transcription system used a start point nearer to box A than with the 16S/23S DPE-sequence. Several other archaeal promoters carrying stretches of thymines in this region also showed a reduction in the distance between box A and the start site. A stretch of more than three thymines causes bending of the DNA which therefore could be the reason for the reduction of this distance.

ACKNOWLEDGEMENTS

We wish to thank Felicitas Pfeifer and Melissa Holmes for critical reading of the manuscript.

REFERENCES

1. Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.
2. Reiter, W-D., Palm, P., Zillig, W. (1988) *Nucl. Acid Res.* **16**, 1–19.
3. Thomm, M. & Wich, G. (1988) *Nucl. Acid Res.* **16**, 151–163.
4. Hüdepohl, U., Reiter, W.-D., Zillig, W.(1990) *Proc. Natl. Acad. Sci. USA*, 87, 5851–5855.
5. Reiter, W-D., Hüdepohl, U., Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9509–9513.
6. Hüdepohl, U., Gropp, F. Horne, M., Zillig, W.(1991) *FEBS Lett*, 285, 257–259
7. Hausner, W., Frey, G. and Thomm, M. (1991) *J. Mol. Biol.* **222**, 495–508
8. Sambrook, J., Fritsch, E.F. and Maniatis, T.: Molecular Cloning, Cold Spring Harbor Laboratory Press, 2nd ed., New York 1989
9. Chung, C., Niemela, S. & Miller, R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2172–2175.
10. Birnboim, H.C. & Doly, J. (1979) *Nucl. Acid Res.* **7**, 1513–1423
11. Sanger, F., Nicklen, S. &Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Murray, M. G. (1986) *Anal. Biochem.* 158, 165–170.
13. Aldea, M., Claverie-Martin, F., Diaz-Torres, M. R. & Kushner, S. R. (1988) *Gene* **65**, 101–110.
14. Reiter, W.-D., Palm, P., Voos, W., Kaniecki, J., Grampp, B., Schulz, W. and Zillig, W. (1987) Nucl. Acid Res. **15**, 5581–5595.
15. Wich, G., Hummel, H., Jarsch, M., Bär, U. and BÖck, A. (1986) *Nucl. Acid Res.* **14**, 2459–2479.
16. Dennis, P.P. (1985) *J. Mol. Biol.* 186, 457–461.
17. Shimmin, L. & Dennis, P.P. (1989) *EMBO J.* **8**, 1225–1235.
18. May, B. & Dennis, B.P. (1990) *J. Bacteriol.* **172**, 3725–3729.
19. Leffers, H., Gropp, F., Lottspeich, F., Zillig, W. and Garrett, R.A. (1989) *J. Mol. Biol.* **206**, 1-17.
20. Blanck, A. & Oesterhelt, D. (1987) *EMBO J.* **6**, 265–273.
21. DasSarma, S., RajBhandary, U.L., Khorana, H.G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 125- 129.
22. Lechner & J., Sumper, M. (1987) *J. Biol. Chem.* **262**, 9724–9729.
23. Gerl, L. & Sumper, M. (1988) *J. Biol. Chem.* **263**, 13246–13251.
24. Salin, M.L., Duke, M.V., Ma, D.-P. and Boyle, J.A. (1991) *Free Rad. Res. Comms.* **12–13**, 443-449.
25. Horne, M., Englert, C. and Pfeifer, F. (1988). *Mol. Gen. Genet.* **213**, 459–464.
26. Jones, J.G., Hackett, N.R., Halladay, J.T., Scothorn, D.J., Yang, C.-F., Ng, W.-L. and Dassarma, S. (1989) *Nucl. Acid Res.* **17**, 7785–7793.
27. Gropp, F., Palm, P. and Zillig, W. (1989) *Can. J. Microbiol.* **35**, 182–188.
28. Gropp, F., Grampp, B., Stolt, P. Palm, P. and Zillig, W. *Virology*, in press
29. Stolt, P. and Zillig, W. submitted 30. Schnabel, H., Palm, P., Dick, K. and Zillig, W. (1984) *EMBO Journal* **3**, 1717–1722.
31. Stolt, P. unpublished
32. Mevarech, M., Hirsch-Twizer, S., Goldman, S., Yakobson, E., Eisenberg H. and Dennis, P.P. (1989) *J. Bacteriol.* **171**, 3479–3485.
33. Reeve, J.N., Beckler, G.S., Cram, D.S., Hamilton, P.T., Brown, J.W., Krzycki, J.A., *Kolodziej, A.F., Alex, L., Orme-Johnson, W.H. and Walsh, C.T. (1989) Proc. Natl. Acad. Sci. USA* **86**, 3031–3035.
34. Bokranz, M., Bäumer, G., Allmansberger, R., Ankel-Fuchs, D. and Klein, A. (1988) *J. Bacteriol.* **170**, 568–577.
35. Brown, J.W. & Reeve, J.N. (1989) FEMS Microbiol. Lett. **60**, 131–136.
36. Auer, J., Spicker, G. and BÖck, A. (1989) *J. Mol. Biol.* **209**, 21–36.
37. Brown, J.W., Thomm, M., Beckler, G.S., Frey, G., Stetter, K.O. and Reeve, J.N. (1988) *Nucl.* Acid Res. **16**, 135–150.
38. Thomm, M., Sherf, B.A. and Reeve, J.N. (1988) *J. Bacteriol.* 170, 1958–1961.
39. Lechner, K., Heller, G. and BÖck, A. (1989) *J. Mol. Evol.* **29**, 20–27.
40. Ree, H. & Zimmermann, R.A. (1990) *Nucl. Acids Res.* **18**, 4471–4478.
41. Klenk, H.-P, Renner, O., Schwass, V. and Zillig, W. *Nucl. Acid Res.* **20**, in press
42. Kletzin, A. submitted
43. Kletzin, A. J. Bacteriol. 174, in press, X56616
44. Kjems, J. & Garrett, R.A. (1987) *EMBO J.* **6**, 3521–3530.
45. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R.A. and *Zillig, W. (1989) Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.
46. Klenk, H.-P., unpublished
47. Auer, J., Spicker, G., Mayerhofer, L., Pühler, G. and BÖck, A. (1991) *System. Appl.* Microbiol. 14, 14–22.
48. SchrÖder, J. & Klink, F. (1991) *Eur. J. Biochem.* **195**, 321–327
49. Reiter, W-D. & Palm, P. (1990) *Mol Gen Genet* **221**: 65–71
50. Culham, D.E. & Nazar, R.N., (1989) *Mol. Gen. Genet.* **216**, 412–416
51. Klenk, H.-P., Palm, P., Lottspeich, F. and Zillig, W. (1992) *Proc. Natl. Acad. Sci. USA* 89, 407–410.
52. Klenk, H.-P., Schwass, V. and Zillig, W. (1991) *Nucl. Acid Res.* **19**, 6047.
53. Kjems, J., Leffers, H., Olesen, T., Holz, I. and Garrett, R.A. (1990) *System. Appl. Microbiol.* **13**, 117–127.
54. Wich, G., Leinfelder, W., and BÖck, A. (1987) *EMBO J.* **6**, 523–528.
55. Betlach, M., Friedman, J., Boyer, H.W. and Pfeifer, F. (1984) *Nucl. Acid Res.* **12**, 7949–7959.
56. Englert, C., Horne, M. and Pfeifer, F. (1990) *Mol. Gen. Genet.* **222**, 225–232.