Mol Gen Genet (1982) 186:339–346

MGG

Nucleotide Sequences that Signal the Initiation of Transcription and Translation in *Bacillus subtilis*

Charles P. Moran Jr., Naomi Lang, Stuart F.J. LeGrice, Gloria Lee, Michael Stephens, A.L. Sonenshein, Janice Pero, and Richard Losick

Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, MA 02138, USA, and The Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA 02111, USA

Summary. We have determined the nucleotide sequence of two *Bacillus subtilis* promoters (*veg* and *tms*) that are utilized by the principal form of *B. subtilis* RNA polymerase found in vegetative cells (σ^{55}-RNA polymerase) and have compared our sequences to those of several previously reported *Bacillus* promoters. Hexanucleotide sequences centered approximately 35 (the "−35" region) and 10 (the "−10" region) base pairs upstream from the *veg* and *tms* transcription startpoints (and separated by 17 base pairs) corresponded closely to the consensus hexanucleotides (TTGACA and TATAAT) attributed to *Escherichia coli* promoters. Conformity to the preferred −35 and −10 sequences may not be sufficient to promote efficient utilization by *B. subtilis* RNA polymerase, however, since three promoters (*veg, tms* and *E. coli tac*) that conform to these sequences and that are utilized efficiently by *E. coli* RNA polymerase were used with highly varied efficiencies by *B. subtilis* RNA polymerase.

We have also analyzed mRNA sequences in DNA located downstream from eight *B. subtilis* chromosomal and phage promoters for nucleotide sequences that might signal the initiation of translation. In accordance with the rules of McLaughlin, Murray and Rabinowitz (1981), we observe mRNA nucleotide sequences with extensive complementarity to the 3′ terminal region of *B. subtilis* 16S rRNA, followed by an initiation codon and an open reading frame.

Introduction

Gene transplantation experiments have shown that *Escherichia coli* is promiscuous in its ability to recognize transcription and translation signals from a wide variety of microorganisms. Two striking examples are the expression in *E. coli* of drug resistance genes from Gram-positive bacteria (Cohen et al. 1973; Ehrlich 1978) and certain amino acid biosynthetic genes from *Saccharomyces cerevisiae* (Struhl et al. 1976; Ratzkin and Carbon 1977). *Bacillus subtilis*, in contrast, seems extremely limited in its ability to express genes from other genera (Kreft et al. 1978); the only clear examples of foreign genes utilized efficiently in *B. subtilis* are drug resistance genes from other Gram-positive bacteria such as *Staphylococcus* (Ehrlich 1978; Kreft et al. 1978) and *Streptococcus* (Yago et al. 1978). Heterologous gene expression is probably restricted at both the transcriptional and translational levels, since *B. subtilis* RNA polymerase fails to transcribe efficiently from *lacUV5* (Lee et al. 1980) and coliphage T4 (Shorenstein and Losick 1973) promoters and to form stable, rapidly starting complexes with coliphage T7 promoters (Wiggs et al. 1979) and *B. subtilis* ribosomes fail to translate effectively from *E. coli* messenger RNAs (Stallcup and Rabinowitz 1973; Legault-Demare and Chambliss 1975). To uncover distinctive features of the transcription and translation initiation signals of *B. subtilis*, we have determined the nucleotide sequences of two *B. subtilis* chromosomal promoters that are utilized by the principal form of *B. subtilis* RNA polymerase found in vegetative cells (σ^{55}-RNA polymerase). [Vegetative and sporulating cells of *B. subtilis* contain additional forms of holoenzyme that differ substantially in their promoter recognition specificity from each other and from σ^{55}-RNA polymerase (Losick and Pero 1981).] We have also searched for possible ribosome binding sites in mRNA sequences in DNA downstream from several chromosomal and phage transcription initiation sites. Combined with previously reported sequences, our results demonstrate striking conservation in the −35 and −10 sequences of *B. subtilis* promoters utilized by σ^{55}-RNA polymerase and, in accordance with the rules of McLaughlin, Murray and Rabinowitz (1981), extensive complementarity to the 3′ terminal region of 16S rRNA in the putative ribosome binding sites of *B. subtilis* mRNAs.

Methods

DNA Sequencing

Nucleotide sequences were determined by the method of Maxam and Gilbert (1980) with DNA strands radioactively labeled at their 5′ termini by the phosphate exchange reaction of phage T4 kinase (Berkner and Folk 1977) or at their 3′ termini by extension with the Klenow fragment of *E. coli* DNA polymerase I (Klenow and Henningson 1970). When necessary, DNA strands were separated by electrophoresis of denatured endonuclease restriction fragments through a 5% polyacrylamide gel (Maxam and Gilbert 1980). The strategy for sequencing individual DNAs are listed below.

veg; the promoter proximal region of the *veg* gene contained on an *Eco*RI-*Hae*III fragment of 530 bp (Ollington

Offprint requests to: R. Losick, The Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA

0026-8925/82/0186/0339/$01.60

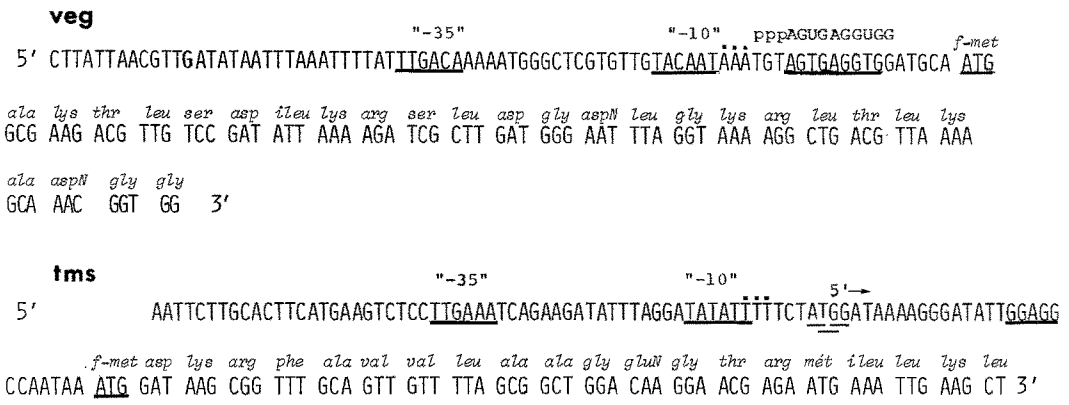

Fig. 1. Nucleotide sequence of the promoter-proximal regions of the *veg* and *tms* genes. The figure shows the nucleotide sequence of the non-transcribed strand of the promoter-proximal region of the *veg* and *tms* genes and summarizes the evidence for the position of the *veg* and *tms* transcription startpoints. The dots above the bases show the *apparent* position of the 5′ termini of run-off RNA as determined in the experiment of Fig. 2; the 5′ RNA sequence above the *veg* DNA sequence is the 5′ terminal sequence of [γ-^{32}P]ATP-labeled *veg* RNA as determined in the experiment of Fig. 3; the underlined bases at the *tms* startpoint correspond to the pattern of dinucleotide transcription priming. Also underlined are the −35 and −10 regions and proposed ribosome binding sites and initiation codons

and Losick 1981) was subcloned into pBR322 by the addition of a *Bam*HI linker at the *Hae*III site, thereby creating plasmid pMS530. The insert in pMS530 was radioactively labeled in separate experiments at either its 5′ or 3′ termini. After strand separation, both strands were sequenced through the promoter region by base-specific cleavage.

tms; a *tms* promoter-containing segment of 700 bp (Ollington et al. 1981) was released from plasmid pLS5 (Donnelly and Sonenshein 1982) by cleavage with *Hin*dIII and labeled separately at either its 5′ or 3′ termini. A 280 bp fragment labeled uniquely at the promoter downstream *Hin*dIII site was then generated by cleavage of the *Hin*dIII-700 fragment at a unique *Eco*RI site. For some experiments, the same 280 bp fragment was released from pLSΔR1 (Donnelly and Sonenshein 1982) and purified by zone sedimentation in a 5–21% gradient of sucrose (dissolved in 10 mM Tris, pH 8 and 1 mM EDTA). Both strands of the *tms* promoter region were sequenced.

SP01 # 15, # 19, # 25; the promoter containing fragments were isolated, labeled, and sequenced as described in Lee and Pero (1981) for SP01 promoters 15, 19.1, and 25.2. To locate the three potential ribosome binding sites the nucleotide sequences of additional DNA downstream from the promoters was determined.

SP01 # 17; a late gene promoter-containing fragment of 480 bp was released from plasmid pMB9–17 (pMB9 containing SP01 *Eco*RI fragment 17 (Lee et al. 1980) by cleavage with *Hin*fI and labeled at its 5′ termini. Subsequent cleavage with *Hin*dIII generated a 320 bp fragment labeled uniquely at the *Hin*fI site downstream from the promoter. The nucleotide sequence of the DNA about 20–200 bp from the *Hin*fI site was determined. (Jolly, Hannett, Costanzo, Lee and Pero, manuscript in preparation.)

RNA Synthesis

Unless otherwise noted, RNA polymerase was prebound to the plasmid DNA template (2 μg) in 40 μl reaction mixtures containing 40 mM Tris, pH 8.0; 10 mM $MgCl_2$; 0.1 mM EDTA; 0.1 mM dithiothreitol; 0.25 mg/ml of bovine serum albumin; and 10% glycerol. RNA synthesis was initiated by the addition of 0.15 mM CTP, GTP, ATP, and 0.4 μM (10 μCi) α^{32}P-UTP. In RNA run-off experiments, heparin was added 60 sec after the addition of ribonucleotides to block reinitiatiton. After 10 min at 37° C, unlabeled UTP (0.15 mM) was added. The reactions were stopped after 5 min by ethanol precipitation and subjected to electrophoresis through polyacrylamide slab gels containing 7 M urea.

Results

Nucleotide Sequence of the veg and tms Promoters

Figure 1 shows the nucleotide sequence (see the Methods) of two previously cloned promoters (*veg* and *tms*) that are recognized by the principal form of *B. subtilis* sigma factor (σ^{55}) found in vegetative cells. Both promoters map within the *purA-cysA* region of the *B. subtilis* chromosome (Haldenwang et al. 1980) at or near sites of transcription initiation in growing cells (Ollington et al. 1981; Donnelly and Sonenshein 1982; Naomi Lang, unpublished results). The location of the *veg* and *tms* promoters in cloned DNA was estimated previously from the sizes of "run-off" RNAs generated by transcription of templates that had been cleaved endonucleolytically within the respective genes (Ollington and Losick 1981; Ollington et al. 1981; Haldenwang and Losick 1980).

To locate the promoters more precisely, we subjected "run-off" RNAs to high resolution gel electrophoresis in lanes next to fragments produced by the base-specific cleavage reactions of Maxam and Gilbert (1980). The DNA fragments were radioactively labeled at the 5′ terminus of the transcribed strand and denatured before electrophoresis. [Since RNA is known to migrate slightly more slowly than does DNA of complementary sequence (Janice Pero, unpublished results; Moran et al. 1981; Moran et al. 1981), the length of the run-off RNAs corresponded only approximately to the *veg* and *tms* startpoints.] Figure 2 shows that the length of the *veg* run-off RNA corresponded to the sequence AAA (indicated by "dots" in Fig. 1) in the non-

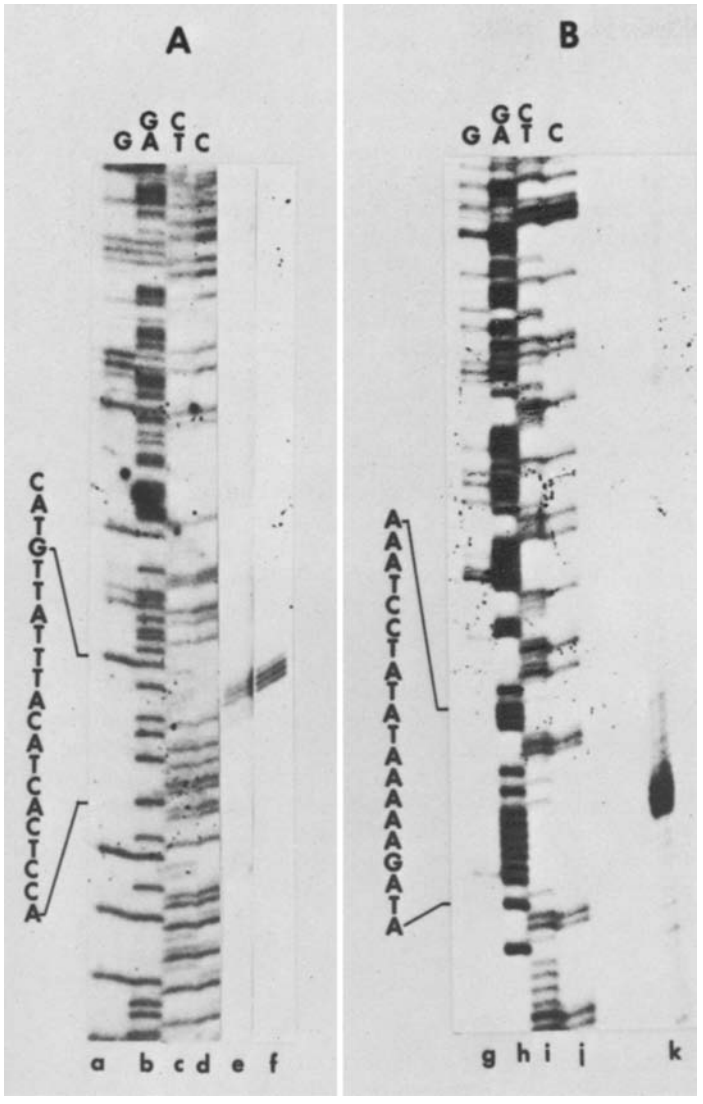

Fig. 2. High resolution gel electrophoresis of *veg* and *tms* run-off transcripts. Panel A; *veg* run-off RNA (tracks e and f) was generated by transcription of the *Bam*HI-530 bp insert from pMS530 and subjected to electrophoresis along side base-specific fragments (tracks a–d) of the transcribed strand of *Bam*HI-530 DNA ^{32}P-labeled at the 5′ terminus (the site of run-off transcription). Panel B; *tms* run-off RNA (track k) was generated by transcription of the *Hin*dIII-700 base pair insert from pLS5 and subjected to electrophoresis along side base-specific fragments (tracks g–j) labeled at the 5′ terminus of the *Hin*dIII run-off transcription site. (DNA uniquely labeled at the run-off transcription site was generated by cleaving *Hin*dIII-700 DNA labeled at both termini with *Eco*R1 and gel-purifying the *tms*-containing *Eco*R1-*Hin*dIII-400 subfragment.) Note that the DNA sequences labeled in panels A and B are of the DNA *coding* strand and thus correspond to the complement of that shown in Fig. 1

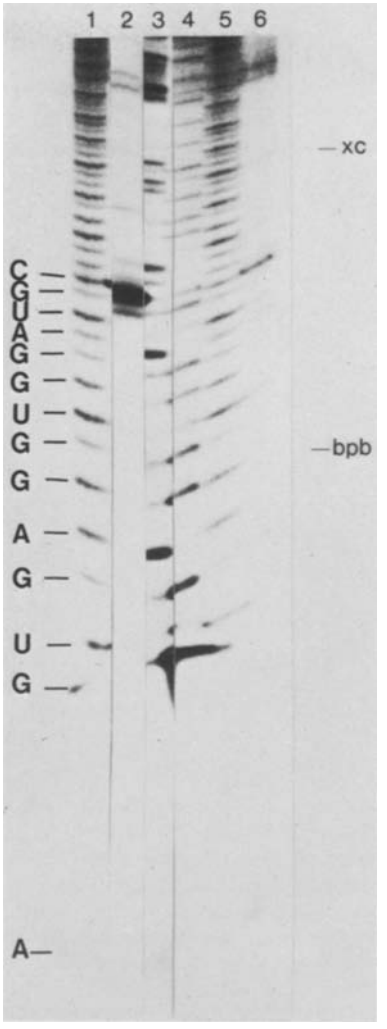

Fig. 3. Sequencing the 5′ terminus of *veg* RNA. *veg* RNA that was uniquely labeled at its 5′ terminus was generated by employing [γ-^{32}P]ATP (100 μM) as a substrate for in vitro transcription in a reaction mixture containing the *Bam*H1-530 insert of pMS530 as template. The resulting end-labeled run-off RNA was gel purified and partially hydrolyzed by alkaline digestion (tracks 1 and 5) or by incubation with RNase CL3 (C-specific; track 2), RNase U2 (A-specific; track 3) or RNase T1 (G-specific; track 4) (Donis-Keller 1979). Both uncleaved RNA (containing a small amount of a stutter transcript; track 6) and the partially digested RNAs (tracks 1–5) were subjected to electrophoresis through a 22% polyacrylamide gel containing 7 M urea. The position of xylene cyanol (XC) and bromophenol blue (bpb) size markers are indicated

transcribed strand (TTT in the transcribed strand); the *tms* run-off transcript migrated at the position of the sequence TTT (AAA in the transcribed strand).

To locate the startpoint of veg transcription directly, we employed [γ-^{32}P]-ATP to label the 5′ terminus of *veg* RNA in an in vitro transcription reaction. The first 14 bases of the end-labeled *veg* RNA were then sequenced by cleavage with base-specific ribonucleases (Donis-Keller 1979) and high resolution gel electrophoresis (Fig. 3). The RNA sequence obtained (pppAGUGA...) is shown in Fig. 1.

Because the *tms* sequence is not transcribed efficiently (see below), we were unable to produce enough 5′-end-labeled transcript to determine the startpoint directly. Therefore, we located the startpoint for *tms* RNA synthesis by examining the pattern of transcription priming by dinucleotides. Dinucleotides are known to prime the initiation of transcription at sites close to or identical to the startpoint of RNA synthesis under conditions of low nucleoside triphosphate concentration (Hoffman and Niyogi 1973; Minkley and Pribnow 1973). Using template DNA that had been cleaved at the *Hin*dIII site within the *tms* transcription unit, we found that AU, UG and GG (but not

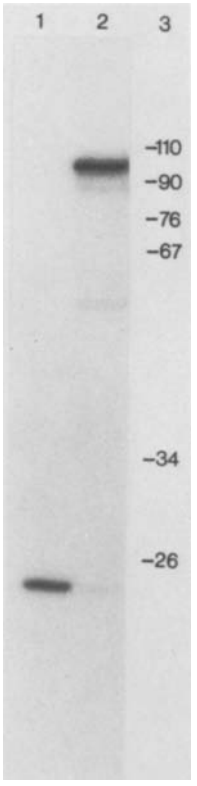

Fig. 4. *tms* RNA synthesized in the absence of CTP. *tms* RNA was generated by transcription of the *Eco*R1-*Hin*dIII-280 insert of pLS5*Δ*R1 in the presence of all four ribonucleoside triphosphates (track 2) or in the absence of CTP (track 1). Reactions conditions were as described under Methods except that RNA polymerase was added to the reaction mixture at 4° C, which was incubated for 1 min at 4° C and then at 37° C for 10 min. The reactions were stopped by addition of an equal volume of Tris-buffered phenol. The aqueous phase was separated, extracted with ether and treated with ethanol to precipitate nucleic acids. Redissolved pellets were subjected to electrophoresis through a 20% polyacrylamide gel containing 7 M urea. The positions of single-stranded DNA size markers (*Hpa*II-cleaved pBR322) are shown in track 3

AA, AC, AG, GA or UA) primed run-off RNA synthesis from the *tms* promoter (data not shown). This corresponds uniquely to the sequence ATGG (as underlined in Fig. 1) whose location in the *tms* promoter sequence is in excellent agreement with the startpoint estimated from the run off transcription experiment of Fig. 2B (remembering that RNA migrates more slowly than does DNA). The initiating nucleotide must be GTP because *tms* RNA synthesis in the absence of dinucleotides required high GTP concentration (150 μM) but required only 5 μM ATP, CTP or UTP. (Since we cannot distinguish which of the two neighboring G residues is the actual startpoint, we have arbitrarily designated the upstream G as the "+1" position.)

As a further check on our proposed *tms* startpoint we transcribed the *tms* gene in the presence of only three nucleoside triphosphates, i.e., ATP, GTP and UTP. Since the first C in the sequence appears 21 bases downstream from the designated startpoint (see Fig. 1), we expected that transcription in the absence of CTP would generate a transcript of approximately 21–22 bases. This expectation was confirmed in the experiment of Fig. 4. A similar sized transcript was also the sole product of a reaction containing *Hae*III-cleaved DNA (the *Hae*III site corresponds to the position of incorporation of the first C residue in *tms* mRNA; data not shown).

In our *tms* transcription experiments, we routinely included only a low concentration of UTP (0.4 μM) in the RNA synthesis reaction mixtures in order to label the product of in vitro transcription to high specific activity. We have discovered, however, that under reaction conditions of high UTP and GTP concentration (150 μM), σ^{55}-RNA polymerase utilized an additional initiation site located approximately 14 bp upstream from the startpoint indicated in Fig. 1. A detailed analysis of this second initiation site will be published elsewhere (SL and ALS, manuscript in preparation).

Efficiency of Promoter Utilization

The experiment of Fig. 5 compares the efficiencies of *B. subtilis* and *E. coli* RNA polymerases to support run-off transcription from *veg*, *tms* (panel A) and an *E. coli* promoter (panel B) *tac* (kindly provided by T. Roberts) which is a hybrid of the *trp* −35 region (TTGACA) and the *lacUV5* −10 region (TATAAT). In our transcription protocol, RNA polymerase was incubated with excess template at 37° in the absence of nucleoside triphosphates (see the Methods). RNA synthesis was then initiated by the addition of substrate, followed by the addition of heparin to prevent re-initiation. (Thus, in this "pre-binding" protocol, we measured the ability of polymerase to utilize promoters to which it has already had an opportunity to bind.) Although *E. coli* RNA polymerase actively transcribed all three templates, *B. subtilis* holoenzyme efficiently utilized only the *veg* template.

Putative Ribosome Binding Sites in B. subtilis mRNA

McLaughlin et al. (1981) have observed that the ribosome binding sites of Gram-positive mRNAs exhibit extensive complementarity to the 3′ region of *B. subtilis* 16S rRNA. To search for such sequences we have determined mRNA sequences in DNA located downstream from the *veg* and *tms* promoters as well as from six other *B. subtilis* chromosomal and phage promoters that are recognized by a variety of bacterial and phage-specified sigma factors (see Methods). These include one early promoter, two middle promoters and one late promoter of phage SP01 as well as two chromosomal sporulation promoters. In each of these presumed mRNA sequences we encountered within 10–150 bases of the 5′ terminus a region of extensive homology to the 3′ terminus of 16S rRNA followed after several bases by an initiation codon and an open reading frame for translation (as underlined in Fig. 1 and 6). Nucleotides that may base pair with rRNA are highlighted with oversize letters in Fig. 6. Figure 6 also lists previously determined translation initiation sequences for α-amylase (Palva et al. 1981), β-lactamase (Kroyer and Chang 1981; Neugebauer et al. 1981) and phage φ29 mRNAs (Murray and Rabinowitz 1982). The predicted free-energy of base-pairing (*Δ*G) of these ribosomal binding sites with 16S rRNA, calculated according to the rules of Tinoco et al. (1973), ranged from −14 to −23 Kcal. To calculate the distances between the ribosome binding sites and their downstream initiation codons, we followed the convention for *E. coli* mRNAs by

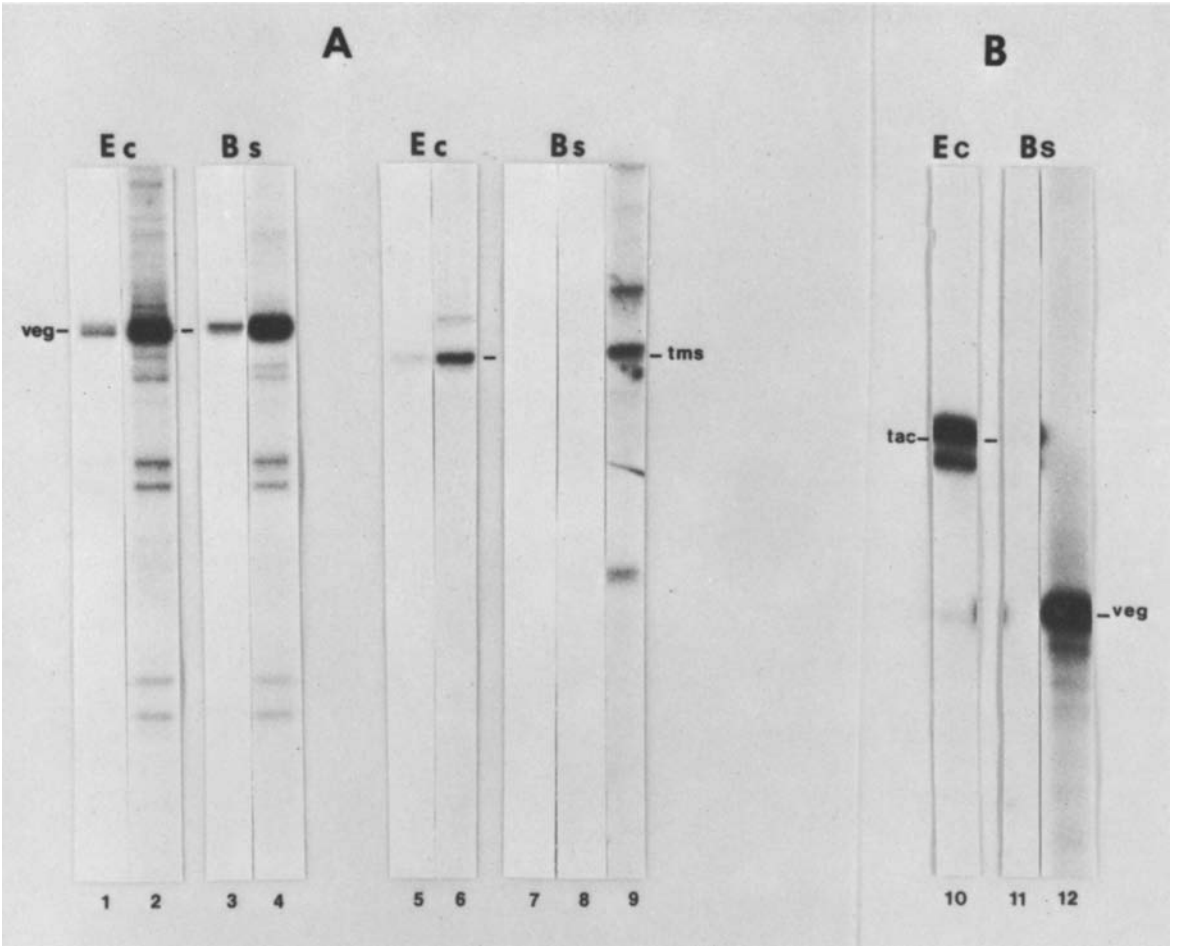

Fig. 5. Efficiencies of utilization of *veg, tms* and *tac* by *B. subtilis* and *E. coli* RNA polymerase. *E. coli* (Ec) RNA polymerase (0.4 units) and *B. subtilis* (Bs) RNA polymerase (0.5 units) were used to transcribe (see the Methods) *Bam*H1-cut pMS530 DNA containing the *veg* promoter, *Hin*dIII-cut pLS5*Δ*R1 containing *tms* promoter, and *Bgl*II-cut DNA containing the *tac* promoter. After electrophoresis through either a 12% (panel A) or a 6% (panel B) polyacrylamide gel containing 7 M urea, the run-off transcripts were visualized by autoradiography for 1,12 or 120 h. Panel A: tracks 1–2, Ec with *veg* DNA, exposed 1 (track 1) and 12 (track 2) h; tracks 3–4 Bs with *veg* DNA, exposed 1 (track 3) and 12 (track 4) h; tracks 5–6, Ec with *tms* DNA, exposed 1 (track 5) and 12 (track 6) h; tracks 7–9, Bs with *tms* DNA, exposed 1 (track 7), 12 (track 8) and 120 (track 9) h. Panel B: track 10, Ec with *tac*; track 11, Bs with *tac*, track 12, Bs with *veg* (tracks 10, 11 and 12 were exposed 12 h)

measuring from the first base to the right of the preferred *E. coli* ribosome binding sequence AGGA (or the equivalent position when this sequence was not present) through the base adjacent to the initiation codon. These distances ranged from 7–14 bases.

Discussion

Promoters

The nucleotide sequences of nine promoters that are utilized by *B. subtilis* σ^{55}-RNA polymerase are listed in Fig. 7. All nine promoters display striking conformity in their −35 and −10 regions with the corresponding consensus nucleotides (TTGACA and TATAAT, respectively) for promoters recognized by *E. coli* RNA polymerase (Rosenberg and Court 1979; Siebenlist et al. 1980). Indeed, bases at five out of twelve positions are invariant in all nine promoters and bases at three additional positions are conserved in eight out of nine promoters (see Fig. 7). Moreover, the "spacer" between the −35 and −10 regions (17 or 18 base pairs) corresponds to the preferred interval for *E. coli* promoters. This conservation of promoter structure seems to be more strict than that observed (Rosenberg and Court 1979) for *E. coli* promoters, (although not enough *Bacillus* promoters have been analyzed to be confident that this will be generally true). Thus, two aspects of these observations are noteworthy: The consensus sequences for the −10 and −35 regions of *B. subtilis* promoters recognized by σ^{55}-RNA polymerase are identical to those suggested for *E. coli* promoters; and, there may be less divergence from these sequences among these *B. subtilis* promoters than there is among the *E. coli* promoters that have been analyzed.

B. subtilis RNA polymerase is known to be less promiscuous than *E. coli* RNA polymerase in the range of promoters with which it interacts strongly (Wiggs et al. 1979; Shorenstein and Losick 1973; Lee et al. 1980). This inability of the *B. subtilis* enzyme to transcribe from certain promoters that are efficiently utilized by *E. coli* RNA polymerase may be explained, in part, by the failure of typical *E. coli* promoters to conform adequately to the canonical nucleo-

mRNA	Nucleotide Sequence	Δ G (kcal)	Spacer (bases)	Position of start codon
veg	PPPAGUGAGGUGGAUGCAAUGGCG	-18	10	+16
tms	AGGGAUAUUGGAGGCCAAUAAAUGGAU	-14	9	+29
spoVG	GGGAAAAGGUGGUGAACUACUGUGGAA	-19	11	+25
0.3 kb	CAGUAAAAAAUGGAGGCUGGGGUGAAA	-17	7	+129
SPO1 #15	AGCGUAAAAGGAGCGGUUAACAUGAUC	-16	9	+53
SPO1 #25	PPPAAAGGAGGAGAGGUUAUUGAGC	-19	10	+17
SPO1 #19	PPPGAAUGGAAGGAGGUAACAAAAUGACC	-20	9	+21
SPO1 #17	UUAUUAUGAGGAGGGAUUUAAAUGCCA	-23	9	+43
ø29 22.4	CAAUCAUAGGAGGAAUUACACAUGAAU	-17	10	+36
ø29 13.9	AAUAGAAAGUGGGACGAAGAAAUGGCA	-18	10	+54
α -amylase	AAAAUGAGAGGGAGAGGAAACAUGAUU	-17	12	+26
β -lactamase	AAACGGAGGGAGACGAUUUUGAUGAAA	-20	14	+159
16S rRNA 3'	UCUUUCCUCCACUAG			

Fig. 6. Nucleotide sequences of putative *Bacillus* ribosome binding sites

	"-35"	"-10"	5'→
veg	AATTTTATTTGACAAAAAT	GGGCTCGTGTTGTACAATAAATGTAGT	
tms	AAGTCTCCTTGAAATCAGA	AGATATTTAGGATATATTTTTCTATGG	
pen	AAAAACGGTTGCATTTAAA	TCTTACATATGTAATACTTTCAAAGAC	
SP01 -15	AAAAGGTATTGACTTTCCC	TACAGGGTGTGTAATAATTTAATTACA	
SP01 -26	AAAAGTTGTTGACTTTATC	TACAAGGTGTGGCATAATAATCTTAAC	
spoOH	GGATTAGGTTGACGCTTTTTTGTCCATTACTGTATAATATTTCATCT		
ϕ29 G3b	GAAAAGTGTTGAAAATTGT	CGAACAGGGTGATATAATAAAAGAGTA	
ϕ29 G2	GAAAAGGGTAGACAAACTATCGTTTAACATGTTATACTATAATAGAA		
ϕ29 A1	ATTAATGTTTGACAACTAT	TACAGAGTATGCTATAATGGTAGTATC	
	TTGACA	TATAAT	
	9 8 9 8 6 5	6 9 8 9 6 9	

Fig. 7. *B. subtilis* promoter nucleotide sequences. The underlined bases show the position of the *veg, tms, penP* (Kroyer and Chang 1981; S. Chang, personal communication) and SP01 early gene promoter (Lee and Pero 1981; Lee et al. 1980) transcription startpoints. The startpoints for the *spoOH* (I. Smith, personal communication) and ϕ29 (Murray and Rabinowitz 1982) promoters have not been precisely determined but were estimated from the sizes of run-off transcription products. An additional ϕ29 promoter (G3a; Murray and Rabinowitz 1982) is not included here because it is utilized very poorly by σ^{55}-RNA polymerase and its startpoint was not determined directly

tides or to their preferred spacing of 17 or 18 base pairs. Thus, *E. coli lacUV5* DNA, which is a poor template for *B. subtilis* RNA polymerase (Lee et al. 1980), lacks the invariant G at the third position of the −35 region. Also, the canonical *E. coli* promoter *tac*, which contains exactly the −35 and −10 prototype sequences, but separated by a spacing of only 16 base pairs, was poorly utilized by *B. subtilis* RNA polymerase (Fig. 5).

Lack of fidelity to the canonical −35 and −10 sequences may not be a sufficient explanation for the failure of *B. subtilis* RNA polymerase to utilize certain promoters, however. Thus, the nearly canonical *tms* promoter, which has a preferred spacer of 17 base pairs and contains all of the most highly conserved −35 and −10 bases, was considerably poorer than *veg* in supporting transcription by σ^{55}-RNA polymerase (although both DNAs were effec-

tively utilized by *E. coli* RNA polymerase; Fig. 5). This could indicate that conformity at the less highly conserved positions (e.g. the fifth positions of the *tms* −35 and −10 hexamers) is more important for promoter recognition by *B. subtilis* RNA polymerase than by *E. coli* RNA polymerase and/or that features in addition to the −10 and −35 hexamers and their spacing are critical in *B. subtilis* promoter utilization.

What additional feature(s) of *B. subtilis* promoters may be important in their efficient utilization by σ^{55}-RNA polymerase? In other work we (Moran et al. 1981) have shown directly by deletion analysis that an AT-rich sequence (25 out of 26 bp) located upstream from the −35 region of the sporulation gene *spoVG* is critical in its utilization by σ^{37}-RNA polymerase. This AT-rich sequence is largely composed of alternating stretches of As and Ts. Similar AT-rich upstream sequences may also promote efficient utilization of σ^{55}-controlled promoters, since such a block (30 bp) precedes the −35 region of the *veg* promoter (see Fig. 1). Alternating stretches of As and Ts are also a feature of the upstream regions of two strong phage SP01 promoters-fragment 15 (AAAAATTTTA-AAAAA) and fragment 26 (AAATT--TT-AAAAATTT--AAAAA) (Lee and Pero 1981; Lee et al. 1980; G. Lee, Ph.D. thesis, Harvard University, 1981) – and three tentatively identified *B. subtilis* rRNA operon promoters (P. Zuber, G. Stewart and K. Bott, personal communication). A test of whether these upstream sequences play a critical role in promoter utilization by σ^{55}-RNA polymerase will require deletion analysis of the *veg* and phage promoters and/or the fusion of AT-rich upstream sequences to the *E. coli tac* and *B. subtilis tms* promoters.

Another region of *B. subtilis* promoters that has been suggested to play a role in promoter strength is the −35 to −10 "spacer" region. Murray and Rabinowitz (1982) have noted homology (ACAGGGTG or AGGTGTGG) in the −35 to −10 spacer region of ϕ29 and SP01 early promoters and have suggested that these spacer sequences may stimulate promoter utilization by *B. subtilis* RNA polymerase. These sequences cannot, however, be a required feature of *B. subtilis* promoters as they are absent in the chromosomal promoters shown in Fig. 7, including the strong *veg* promoter. We do note, however, that seven of the nine promoters listed in Fig. 7, including the strongest promoters, conform significantly to a less specific sequence PuTPuTG at positions −18 to −14. Such a sequence appears very rarely in *E. coli* promoters (Rosenberg and Court 1979). We have no information on whether this sequence is functionally important in *B. subtilis*.

What features of *B. subtilis* RNA polymerase account for its restricted ability to utilize heterologous promoters? In addition to core and sigma factor, the *B. subtilis* holoenzyme is associated with a 21,000 dalton polypeptide known as delta that enhances the specificity of transcription (Pero, Nelson, and Fox 1975; Tjian et al. 1977; Achberger and Whiteley 1981). The low activity of *B. subtilis* holoenzyme on coliphage T7 DNA has been ascribed to delta (Achberger and Whiteley 1981) whereas the inefficient utilization of coliphage T4 DNA has been attributed to the core (Shorenstein and Losick 1973). It would be interesting to determine whether core or delta accounts for the varied efficiencies of utilization of the *veg*, *tms* and *tac* promoters observed here.

Ribosome Binding Sites

McLaughlin et al. (1981) have proposed that the ribosome binding (Shine-Delgarno) sites of Gram-positive mRNAs are characteristically able to form highly stable complexes with the 3′ terminal region of 16S rRNA, whereas the potential for base pairing by *E. coli* Shine-Delgarno sites varies over a wide range. The ribosome binding sites in four Gram-positive bacterial and phage mRNAs for which the initiation codon has been identified from amino acid sequence information conform well to this proposal [α-amylase (Palva et al. 1981), β-lactamase (Kroyer and Chang 1981; Neugebauer et al. 1981) and the ϕ29 mRNAs (Murray and Rabinowitz 1982) listed in Fig. 6]. As a further test of this proposal we have searched for possible ribosome binding sites in mRNA sequences located downstream from a variety of *B. subtilis* chromosomal and phage promoters including vegetative and sporulation promoters and phage SP01 early, middle and late promoters. Although we have not identified initiation codons directly by amino acid sequencing, all eight mRNAs display regions of extensive complementarity to the 3′ region of *B. subtilis* 16S rRNA followed by a possible initiation codon and an open reading frame. In several cases this complementarity included the extreme 3′ terminus of *B. subtilis* 16S RNA that is nonhomologous to *E. coli* 16S rRNA (UCUUUCCUCC in *B. subtilis* vs. AUUCCUCC in *E. coli*). The calculated free-energies of interaction (ΔG) for our proposed ribosome binding sites vary from −14 to −23 Kcal as compared to −9.4 Kcal for the prototype *E. coli* ribosome binding sequence AGGA, confirming the prediction of McLaughlin et al. (1981). Nevertheless, the distance between ribosome binding sites and initiation codons was not inconsistent with that observed for *E. coli* mRNAs: as measured from the first base to the right of AGGA or its equivalent (in accordance with the convention for *E. coli* mRNAs) through the adjacent base to the initation codon, the distance between ribosome binding sites and initiation codons ranged from 7–14 bases.

Heterologous Gene Expression

In summary, the inability of *B. subtilis* to express genes from *E. coli* and other Gram-negative bacteria (Ehrlich 1978; Kreft et al. 1978) can be understood if *B. subtilis* is more stringent than *E. coli* in its requirements for recognition of transcription and translation initiation signals. Thus the *B. subtilis* σ^{55}-RNA polymerase may demand high fidelity to the canonical −35 and −10 promoter hexanucleotides (and perhaps other sequences as well) while the *B. subtilis* ribosome may require extensive complementarity to messenger RNA. While it has not been proved that the extent of homology to consensus sequences is directly related to initiation efficiency, it will be interesting to determine whether such stringent adherence to transcription and translation initiation signals proves to be a general feature of gene expression in Gram-positive bacteria (Murray and Rabinowitz 1982).

Acknowledgements. This work was supported by a gift from the Eli Lilly Corporation to R. Losick, and research grants from the National Institutes of Health to R. Losick (GM18568), to A.L. Sonenshein (GM19168), and from the National Science Foundation (PCM-7821570) to J. Pero and A.L. Sonenshein (PCM79-00944). We thank T. Roberts for the gift of *tac* DNA.

References

Achberger EC, Whiteley HR (1981) The role of the δ peptide of the *Bacillus subtilis* RNA polymerase in promoter selection. J Biol Chem 256:7424–7432

Berkner KL, Folk WR (1977) Polynucleotide kinase exchange reaction: quantitative assay for restriction endonuclease-generated 5′-phosphoryl termini in DNAs. J Biol Chem 252:3176–3184

Cohen SN, Chang ACY, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids *in vitro*. Proc Natl Acad Sci USA 70:3240–3244

Donnelly CE, Sonenshein AL (1982) Genetic fusion of *Escherichia coli lac* genes to a *Bacillus subtilis* promoter. In: Ganeson A, Hoch J, Chang S (eds) Molecular cloning and gene regulation in *Bacilli*. Academic Press, New York, in press

Donis-Keller H (1979) Site specific cleavage of RNA. Nucl Acids Res 7:179–192

Ehrlich SD (1978) DNA cloning in *Bacillus subtilis*. Proc Natl Acad Sci USA 75:1433–1436

Haldenwang WG, Banner CDB, Ollington JF, Losick R, Hoch JA, O'Connor MB, Sonenshein AL (1980) Mapping a cloned gene under sporulation control by insertion of a drug-resistance marker into the *Bacillus subtilis* chromosome. J Bacteriol 142:90–98

Haldenwang WG, Losick R (1980) A novel RNA polymerase sigma factor from *Bacillus subtilis*. Proc Natl Acad Sci USA 77:7000–7004

Hoffman DJ, Niyogi SK (1973) RNA initiation with dinucleoside monophosphates during transcription of bacteriophage T4 DNA with RNA polymerase of *Escherichia coli*. Proc Natl Acad Sci USA 70:574–578

Kreft J, Bernhard K, Goebel W (1978) Recombinant plasmids capable of replication in *B. subtilis* and *E. coli*. Mol Gen Genet 162:59–67

Klenow H, Henningson I (1970) Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. Proc Natl Acad Sci USA 65:168–175

Kroyer J, Chang S (1981) The promoter-proximal region of the *Bacillus licheniformis* penicillinase gene: Nucleotide sequence and predicted leader peptide sequence. Gene 15:343–347

Lee G, Hannett NM, Korman A, Pero J (1980) Transcription of cloned DNA from *Bacillus subtilis* phage SP01: requirement for hydroxymethyluracil-containing DNA by phage-modified RNA polymerase. J Mol Biol 139:407–422

Lee G, Pero J (1981) Conserved nucleotide sequences in temporally-controlled phage promoters. J Mol Biol 152:247–265

Lee G, Talkington C, Pero J (1980) Nucleotide sequences of a promoter recognized by *Bacillus subtilis* RNA polymerase. Mol Gen Genet 180:57–65

Legault-Demare L, Chambliss GH (1975) Selective messenger translation of *Bacillus subtilis* ribosomes. Mol Gen Genet 142:277–287

Losick R, Pero J (1981) Cascades of sigma factors. Cell 25:582–584

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base specific chemical cleavages. Meth Enzymol 65:499–559

McLaughlin JR, Murray CL, Rabinowitz JC (1981) Unique features in the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus* β-lactamase gene. J Biol Chem 256:11283–11291

Minkley EG, Pribnow D (1973) Transcription of the early region of bacteriophage T7: selective initiation with dinucleotides. J Mol Biol 77:255–277

Moran CP Jr, Lang N, Banner CDB, Haldenwang WG, Losick R (1981) Promoter for a developmentally-regulated gene in *Bacillus subtilis*. Cell 25:783–791

Moran CP Jr, Lang N, Losick R (1981) Nucleotide sequence of a *Bacillus subtilis* promoter recognized by *Bacillus subtilis* RNA polymerase containing σ^{37}. Nucl Acids Res 9:5979–5990

Murray CL, Rabinowitz JC (1982) Nucleotide sequences of transcription and translation initiation regions in *Bacillus* phage ϕ29 early genes. J Biol Chem 257:1053–1062

Neugebauer K, Sprengel R, Schiller H (1981) Penicillinase from *Bacillus licheniformas*: nucleotide sequence of the gene and implications for the biosynthesis of a secretory protein in a Gram-positive bacterium. Nucl Acids Res 9:2577–2588

Ollington JF, Losick R (1981) A cloned gene that is turned on at an intermediate stage of spore formation in *Bacillus subtilis*. J Bacteriol 147:443–451

Ollington JF, Haldenwang WG, Huynh TV, Losick R (1981) Developmentally-regulated transcription in a cloned segment of the *Bacillus subtilis* chromosome. J Bacteriol 147:432–442

Palva I, Pettersson RF, Kalkkinen N, Lehtovaara P, Sarvas M, Soderlund H, Takkinen K, Kaariainen L (1981) Nucleotide sequence of the promoter and NH_2-terminal signal peptide region of the α-amylase gene from *Bacillus amyloliquefaciens*. Gene 15:43–51

Pero J, Nelson J, Fox TD (1975) Highly asymmetric transcription by RNA polymerase containing phage-SP01-induced polypeptides and a new host protein. Proc Natl Acad Sci USA 72:1589–1593

Ratzkin B, Carbon J (1977) Functional expression of cloned yeast DNA in *Escherichia coli*. Proc Natl Acad Sci USA 74:487–491

Rosenberg M, Court D (1979) Regulatory sequences involved in the promotion and termination of transcription. Annu Rev Genet 13:319–353

Shorenstein RG, Losick R (1973) Comparative size and properties of the sigma subunits of RNA polymerase from *Bacillus subtilis* and *Escherichia coli*. J Biol Chem 248:6170–6173

Siebenlist U, Simpson RB, Gilbert W (1980) *E. coli* RNA polymerase interacts homologously with two different promoters. Cell 20:269–281

Stallcup MR, Rabinowitz JC (1973) Initiation of protein synthesis *in vitro* by a clostridial system. J Biol Chem 248:3209–3215; 3216–3219

Struhl K, Cameron JR, Davis RW (1976) Functional genetic expression of eukaryotic DNA in *Escherichia coli*. Proc Natl Acad Sci USA 73:1471–1475

Tinoco I, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM, Gralla J (1973) Improved estimation of secondary structure in ribonucleic acids. Nature New Biol 246:40–41

Tjian R, Losick R, Pero J, Hinnebush A (1977) Pruification and comparative properties of the delta and sigma subunits of RNA polymerase from *Bacillus subtilis*. Eur Journal Biochem 74:149–154

Wiggs JL, Bush JW, Chamberlin MJ (1979) Utilization of promoter and terminator sites on bacteriophage T7 DNA by RNA polymerases from a variety of bacterial orders. Cell 16:97–109

Yago Y, McLellan T, Frez W, Clewell D (1978) Antimicrobial agents and chemotherapy 13:884–887

Communicated by E. Bautz

Received April 14, 1982