Nucleotide sequences of two *Bacillus subtilis* promoters used by *Bacillus subtilis* sigma-28 RNA polymerase

Michael Z. Gilman, Janey L. Wiggs and Michael J. Chamberlin

Department of Biochemistry, University of California, Berkeley, CA 94720, USA

Received 24 July 1981

ABSTRACT

RNA polymerase holoenzymes from many bacterial species share a common promoter recognition specificity since they use the same promoter sites on a variety of templates. These promoters generally include sequences homologous to the average sequences -TTGACA- and -TATAATA-, located -35 and -10 base pairs, respectively, upstream of the transcriptional start site. We have isolated a minor form of *B. subtilis* RNA polymerase in which the normal sigma subunit (σ^{55}) is replaced by a smaller polypeptide (σ^{28}) and which is highly specific for a class of promoter sites not used by the σ^{55} -holoenzyme. Sequencing of two *B. subtilis* promoter sites used by the σ^{28} -holoenzyme reveals identical sequences at -35 and -10 base pairs from the start site, which are -CTAAA- and -CCGATAT-, respectively. These results confirm that sigma subunit plays a major direct role in promoter sequence recognition, and support a model in which sigma interacts sequentially with -35 and -10 regions, respectively.

INTRODUCTION

The sigma subunit of bacterial RNA polymerase is required for normal recognition of promoter sequences in DNA and for initiation of transcription from those sites (1). Recent studies have shown that multiple protein factors exist in both phage infected (2-5) and normal *B. subtilis* (6-10) that can replace the predominant *B. subtilis* sigma subunit (σ^{55}). The resulting RNA polymerases can direct transcription from promoters not used by the σ^{55} -holoenzyme. The DNA sequences of these different classes of promoters should provide important information as to the role of sigma factors in promoter recognition and gene regulation (11,12). We report here the nucleotide sequences of two *B. subtilis* promoter sites used exclusively by a form of *B. subtilis* RNA polymerase (σ^{28} -holoenzyme) in which a polypeptide of Mr 28000 (σ^{28}) replaces σ^{55} (10).

MATERIALS AND METHODS

1. Enzymes. *B. subtilis* RNA polymerases were purified from exponen-

tially growing *B. subtilis* W168 as described (10). For some experiments, σ^{28} -holoenzyme was further purified by chromatography on single-stranded DNA-cellulose (Wiggs and Chamberlin, manuscript in preparation).

Restriction endonucleases were purchased from Bethesda Research Labs (Bethesda, MD), except Eco RI and Hind III, purified in this laboratory. T4 polynucleotide kinase was purchased from P-L Biochemicals (Madison, WI). Calf intestine alkaline phosphatase was a gift of C. Murray and J. Rabinowitz.

2. Construction of pMG102 and pMG201. Plasmids pCD4136 and pCD4322 were isolated from a pooled *B. subtilis* genomic clone library provided by Drs. E. Ferrari and J. Hoch (Scripps Institute). They contain *B. subtilis* genomic fragments bearing promoters for σ^{28} -holoenzyme cloned into vector pHV14, a hybrid of pBR322 and *S. aureus* plasmid pC194 (10). To facilitate sequence analysis, sub-fragments of these original isolates were cloned in pBR322.

pCD4136 was digested with Eco RI, generating three small fragments and one large one containing most of the vector and the leftmost 1.1 kb of insert DNA. The mixture of fragments was ligated at low DNA concentration (5 μ g/ml) to optimize for re-circularization of the large fragment. After transformation into *E. coli* 294, selection for chloramphenicol resistance and screening for plasmid size (13), plasmid pMG101 was recovered. pMG101 retains pC194 DNA, and these sequences were deleted by digestion with Hind III followed by ligation at low DNA concentration. Re-joining of the Hind III ends derived from pBR322 reconstructs the tetracycline resistance gene (promoter); thus, transformants were selected with tetracycline. The resulting plasmid pMG102 (Figure 1) contains 1.1 kb of the original *B. subtilis* DNA insert and most of pBR322.

pCD4322 DNA was cut to completion with Hind III. The 1.6 kb promoter-containing fragment was isolated (14) and ligated to pBR322 DNA linearized by digestion with Hind III. After transformation and selection for ampicillin resistance, recombinant plasmids were visualized by agarose gel electrophoresis (13). Plasmids of appropriate mobility were analyzed by restriction digestion to determine the identity and orientation of the inserts; pMG201 (Figure 1) was recovered.

3. DNA Sequencing. Plasmid DNAs were prepared by a modification of a published procedure (15). DNA to be end-labeled was purified free of low molecular weight contaminants by equilibrium sedimentation in CsCl/ethidium bromide followed by chromatography on Sepharose CL-2B or a second equilibrium sedimentation. Restriction fragments for sequence analysis were 5'-end-labeled with T4 polynucleotide kinase and γ -[32 P]-ATP (16) under exchange con-

ditions (17) or by dephosphorylation with calf intestine alkaline phosphatase and direct phosphorylation (18). Singly-labeled restriction fragments and separated strands were isolated as described (18).

Sequencing reactions were carried out as described by Maxam and Gilbert (18). Cleavage reactions used were: G only (DMS), C only (hydrazine), C+T (hydrazine), and either A+C (NaOH) or A+G (formic acid). Sequencing runs were as follows (numbers in parentheses are positions in base pairs of restriction sites on the insert fragments; see Fig. 1). pMG102: Hinf I (800) leftward and rightward, Alu I (920) leftward, Cfo I (700) rightward. pMG201: Ava I (920) rightward, Hpa II (810) rightward, Hpa II (1000) leftward, Alu I (850) rightward. Roughly 200 nucleotides of sequence were determined from each run.

4. RNA sequencing. γ -[32 P]-labeled RNA for sequence analysis was synthesized in vitro using σ^{28} -holoenzyme and pMG102 and pMG201 DNAs, which had been digested with Hpa II and Cfo I, respectively. Incubations (200 μ l total volume) contained 50 μ M γ -[32 P]-ATP (2.2×10^5 cpm/pmol), 200 μ M each CTP, GTP and UTP, and ionic conditions described previously (10). After 60 min at 37 $^{\circ}$, [32 P]-RNA was isolated by precipitation with ethanol (four times) and was treated with ribonucleases T1 and U2 as described by Donis-Keller *et al.* (19). Digestion products were analyzed by electrophoresis on 20% polyacrylamide/8.3 M urea gels (0.3 x 400 mm).

FIGURE 1. *B. subtilis* chromosomal fragments subcloned in pBR322. Location and orientation of σ^{28} -holoenzyme promoters are indicated. Restriction sites are abbreviated as follows: Al, Alu I; Av, Ava I; C, Cfo I; Hc, Hinc II; Hf, Hinf I; Hp, Hpa II; S, Sau3A I; T, Taq I.

RESULTS

B. subtilis σ^{28} holoenzyme was first detected in *B. subtilis* RNA polymerase preparations as an activity able to initiate transcription from an *in vitro* promoter site near the right end of T7 DNA which is not used by any other bacterial RNA polymerase (9). The σ^{28} holoenzyme has been separated from σ^{55} holoenzyme and shown to exhibit a high degree of template specificity. It shows little or no activity with *E. coli* DNA or DNA from several plasmid vectors (ref. 10 and Table I). It is very active with *B. subtilis* DNA and with certain plasmids containing *B. subtilis* DNA fragments. This is due to the presence of strong promoter sites for σ^{28} holoenzyme on these templates and has permitted the transcriptional mapping of σ^{28} -specific promoters on two *B. subtilis* DNA inserts cloned into vector pHV14 (10).

To determine the nucleotide sequences of these two σ^{28} -specific promoters, we prepared smaller subclones of the original inserts, generating plasmids pMG102 and pMG201 (Fig. 1). Sequences of both DNA strands in the promoter regions of these plasmids (Fig. 2) were determined by the method of Maxam and Gilbert (18).

The locations of the *in vitro* transcriptional start sites on the plasmids were initially estimated by *in vitro* transcription of restriction endonuclease-cleaved templates. Run-off transcripts of 260 and 185 bases were

TABLE I. RNA polymerase assays were carried out as described (10). Plasmid DNA templates were linearized by cleavage restriction endonuclease Pst I, which cuts once in each plasmid except pCD4134, which is cut at two sites.

Activities of <i>B. subtilis</i> RNA Polymerases on Different DNA Templates		
DNA	Holoenzyme Fraction	
	$E\sigma^{55}$ (u/mg)	$E\sigma^{28}$ (u/mg)
<i>B. subtilis</i>	2900	1700
<i>E. coli</i>	800	70
ϕ 29	7800	140
T7	3400	200
pBR322	1100	30
pHV14	2600	70
pCD4134	1400	170
pMG102	2500	4400
pMG201	1900	4800

FIGURE 2. Nucleotide sequences of σ^{28} holoenzyme promoters and RNA transcripts on pMG102 and pMG201. Homologous DNA sequences are underlined. RNA sequences are displayed under each DNA sequence. In the RNA sequences, Y indicates pyrimidine and N is an undetermined base.

obtained by transcription of pMG102 DNA cleaved with Eco RI and Hpa II, respectively (see Fig. 1). Run-off transcripts of 360, 220, and 120 bases were synthesized from pMG201 DNA cleaved with Hinc II, Taq I, and Cfo I, respectively. Cleavage of pMG102 with Hinf I and pMG201 with Hpa II abolished detectable transcription, suggesting that these enzymes cut at or near the promoters.

The initiating nucleotide of each transcript was determined by *in vitro* transcription in the presence of γ -[32 P]-nucleoside triphosphates; γ -[32 P]-ATP, GTP, and CTP were tested. With both plasmids, only γ -[32 P]-ATP gave rise to labeled σ^{28} -specific transcripts.

Finally, the 5'-terminal sequences of the transcripts were ascertained by enzymatic cleavage of γ -[32 P]-ATP-labeled *in vitro* transcripts (19), followed by gel analysis of the cleavage products (Fig. 3). This allowed us to position precisely the transcriptional start sites in each case; these are designated as +1 in Figure 2. Although there are several A's in the vicinity of both start sites, the limit T1 and U2 digests are homogeneous (Fig. 3, tracks 1,3). Thus, all of the labeled RNA transcribed from each plasmid arises from a unique site. The initial RNA samples, untreated with nuclease (track 8), contain small amounts of several short oligonucleotides. We attribute these species to premature termination or pausing of transcription initiated at the specific σ^{28} promoters, since these sequences all share a common T1 and U2 limit 5' terminus. Similar short paused transcripts are well known to occur in *in vitro* transcription products (20,21). Where these oligonucleotides interfere with unambiguous reading of the RNA sequence, we have scored those positions as N (Fig. 2).

Bacterial RNA polymerases normally interact with DNA sequences located between about -40 and +20 base pairs around the start site (22,23). In particular, there are two regions centered around -35bp and -10bp which show a

FIGURE 3. Sequence analysis of σ^{28} -specific RNA transcripts of pMG102 (left) and pMG201 (right) DNAs. Synthesis and isolation of γ -[32 P]-labeled RNA was carried out as described in Materials and Methods. Treatment of RNA with ribonucleases T1 and U2 and alkali digestion were done as described by Donis-Keller et al. (19). Cleavage products were analyzed by electrophoresis on 20% polyacrylamide/8.3 M urea gels. RNA samples in each track were treated as follows: track 1, 500 u/ml T1; track 2, 50 u/ml T1; track 3, 400 u/ml U2; track 4, alkali digestion; track 5, 5 u/ml T1; track 6, 40 u/ml U2; track 7, alkali digestion; track 8, untreated; track 9, 0.5 u/ml T1; track 10, 4 u/ml U2.

high degree of sequence homology for nearly all the promoters used by the predominant forms of bacterial RNA polymerase. These have average sequences of -TTGACA- and -TATAATA-, respectively, for the *E. coli* holoenzyme, and many similar homologies are found for representative *B. subtilis* promoters (24). These regions are critical for promoter utilization or recognition, since pro-

motor mutations are found to alter these sequences almost exclusively. It has been proposed that sigma interacts specifically at one of these regions while the other is recognized by a site on the core RNA polymerase (25).

The two σ^{28} -specific promoter sequences are identical from -35 to -31 (-CTAAA-) and from -14 to -8 (-CCGATAT-) and are considerably different from the average *E. coli* sequences. In addition, there is a 4bp identical sequence at the start site (-1 to +3bp), -TAGA-.

DISCUSSION

The two σ^{28} promoters that we have sequenced in these studies are strong *in vitro* promoters that require only σ^{28} and *B. subtilis* core RNA polymerase for *in vitro* transcription (10). These σ^{28} promoters are not used detectably by the σ^{55} holoenzyme under any *in vitro* conditions we have tested, nor does the σ^{28} holoenzyme use any of the σ^{55} promoters we have studied (9,10). At the present time we know nothing about the function of these promoters or the genes they control *in vivo*. However, recent experiments using the Southern transfer procedure adapted for RNA (26) have shown the σ^{28} promoter cloned in pMG201 is used efficiently *in vivo* (M. Gilman, unpublished observations). The *in vivo* transcript from this promoter is identical in size to the *in vitro* transcript obtained with the parent plasmid pCD4322, which contains a strong termination site about 1600 base pairs downstream from the σ^{28} promoter (10).

There are now at least six factors known that interact with *B. subtilis* core polymerase to confer the ability to use distinct classes of promoter sites (12,27; for review, see ref. 28). In general the promoter sites used by a particular sigma factor are not used by any other sigma, although σ^{37} and σ^{29} overlap in their specificity. In the four cases where promoters used by a particular sigma have been compared, the promoters share homologous or identical sequences at both the -35 and -10 regions. In addition, the actual sequences found in these regions differ substantially from the corresponding sequences of promoters used by a heterologous sigma.

These observations are most simply explained if sigma factors are sequence-specific DNA binding proteins that interact directly with bases at -35 and -10 to allow binding and initiation of RNA polymerase at cognate promoter sites. Formation of RNA polymerase-promoter complexes proceeds in several steps (32) and involves intermediate formation of a complex between the DNA duplex and the polymerase (closed promoter complex) followed by isomerization to a complex in which a limited number of base pairs are open at the start site (open promoter complex). A plausible model would be that a

binding site (site I) on sigma interacts initially with the base pairs in the -35 region to allow closed promoter complex formation, and then is translocated to the -10 region, where a second binding site (site II) binds to the DNA bases of the non-transcribed strand of -10 region to stabilize the open promoter complex (27). The core polymerase must play a significant role in facilitating and stabilizing these interactions, of course. Sigma alone binds to DNA, but not tightly (30,31), and the region of DNA that interacts with RNA polymerase extends over at least 60bp (23), a larger area than would be covered by sigma alone.

Although there is yet no direct evidence for the model, it has been evident for some time that there must be both specific antimeeting and meeting interactions involved in promoter binding, and these would require sites with properties ascribed to sigma sites I and II (29). *E. coli* sigma (σ^{70}) can be crosslinked to BrU at -3 in the *lac* UV5 promoter in open promoter complexes (29) and also can be coupled to the 5' terminus of a dinucleotide primer (33). This is good evidence for the specific binding of sigma near the -10 region. There is also a recent report that *E. coli* sigma can be crosslinked to the -35 region of a partially depurinated *lac* UV5 promoter (31). This provides provocative evidence for the postulated sigma interaction at -35, although the nature of the complex involved is not yet clear.

An alternative form of the model involves sigma making simultaneous contact with both -35 and -10 regions. This cannot be ruled out, but seems less likely in view of the long distances (30 bp, ca 100 Å) and the small size of σ^{28} . More complex possibilities, such as binding of two molecules of σ^{28} to the core polymerase, are also not yet excluded.

We showed several years ago that the predominant forms of RNA polymerase in a variety of bacterial species all share a common promoter recognition identity in that they all transcribe the same promoter sites on a single template (35). This is explained if the predominant species of sigma subunits of different bacteria have similar promoter sequence recognition sites, even though they are of different sizes. However, there are large differences in rates of utilization of particular promoters by heterologous RNA polymerases, and these are not determined by the identity of the sigma subunits but by the core polymerase involved (36,37). This suggests that there may also be important, sequence determined, interactions between core polymerase and promoter regions that affect the efficiency of promoter utilization.

ACKNOWLEDGEMENTS

This research was supported by a research grant (GM 12010) from the National Institutes of Health General Medical Sciences and by a National Institutes of Health General Medical Sciences training grant (GM 07232).

REFERENCES

1. Losick, R. and Pero, J. (1976) in RNA Polymerase, Losick, R. and Chamberlin, M., Eds., pp. 227-246, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
2. Duffy, J. J. and Geiduschek, E. P. (1977) *Nature* 270, 28-32.
3. Tjian, R. T. and Pero, J. (1976) *Nature* 262, 753-757.
4. Talkington, C. and Pero, J. (1978) *Proc. Nat. Acad. Sci. USA* 75, 1185-1189.
5. Whiteley, H. R., Spiegelman, G. B., Lawrie, J. M. and Hiatt, W. R. (1976) in RNA Polymerase, Losick, R. and Chamberlin, M., Eds., pp. 587-600, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
6. Haldenwang, W. G. and Losick, R. (1979) *Nature* 282, 256-260.
7. Haldenwang, W. G. and Losick, R. (1980) *Proc. Nat. Acad. Sci. USA* 77, 7000-7004.
8. Haldenwang, W. G., Lang, N. and Losick, R. (1981) *Cell* 23, 615-624.
9. Jaehning, J. A., Wiggs, J. L. and Chamberlin, M. J. (1979) *Proc. Nat. Acad. Sci. USA* 76, 5470-5474.
10. Wiggs, J. L., Gilman, M. Z. and Chamberlin, M. J. (1981) *Proc. Nat. Acad. Sci. USA* 78, 2762-2766.
11. Talkington C. and Pero J. (1979) *Proc. Nat. Acad. Sci. USA* 76, 5465-5469.
12. Moran, C. P., Jr., Lang, N. and Losick, R. (1981) *Nucleic Acids Res.*, accompanying paper.
13. Barnes, W. M. (1977) *Science* 195, 393-394.
14. Yang, R. C.-A., Lis, J. and Wu, R. (1979) in *Methods in Enzymology*, Wu, R., Ed., Vol. 68, pp. 176-182, Academic Press, New York.
15. Hirt, B. (1967) *J. Mol. Biol.* 26, 365-369.
16. Johnson, R. A. and Walseth, T. F. (1979) in *Advances in Cyclic Nucleotide Research*, Brooker, G., Greengard, P. and Robison, G. A., Eds., Vol. 10, pp. 135-167, Raven Press, New York.
17. Berkner, K. L. and Folk, W. R. (1977) *J. Biol. Chem.* 252, 3176-3184.
18. Maxam, A. W. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K., Eds., Vol. 65, pp. 499-560, Academic Press, New York.
19. Donis-Keller, H., Maxam, A. M. and Gilbert W. (1977) *Nucleic Acids Res.* 4, 2527-2538.
20. Maizels, N. M. (1973) *Proc. Nat. Acad. Sci. USA* 70, 3585-3589.
21. Kasavetis, G. A. and Chamberlin, M. J. (1981) *J. Biol. Chem.* 256, 2777-2786.
22. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genet.* 13, 319-353.
23. Siebenlist, U., Simpson, R. B. and Gilbert W. (1980) *Cell* 20, 269-281.
24. Lee, G., Talkington, C. and Pero J. (1980) *Molec. Gen. Genet.* 180, 57-65.
25. Gilbert, W. (1976) in RNA Polymerase, Losick, R. and Chamberlin, M., Eds., pp. 193-206, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
26. Thomas, P. S. (1980) *Proc. Nat. Acad. Sci. USA* 77, 5201-5205.
27. Lee, G. and Pero, J. (1981) *J. Mol. Biol.*, in press.
28. Losick, R. and Pero, J. (1981) *Cell*, in press.
29. Chamberlin, M. J. (1974) *Ann. Rev. Biochem.* 43, 721-775.
30. Stender, W. (1980) *Nucleic Acids Res.* 8, 3459-3466.

31. Kudo, T., Jaffe, D. and Doi, R. H. (1981) *Molec. Gen. Genet.* 181, 63-68.
32. Simpson, R. B. (1979) *Cell* 18, 277-285.
33. DeRiemer, L. H. and Meares, C. F. (1981) *Biochem.* 20, 1612-1617.
34. Chenchick, A., Beabealashvilli, R. and Mirzabekov, A. (1981) *FEBS Lett.* 128, 46-50.
35. Wiggs, J. L., Bush, J. W. and Chamberlin, M. J. (1979) *Cell* 16, 97-109.
36. Davison, B. L., Murray, C. L. and Rabinowitz, J. C. (1980) *J. Biol. Chem.* 255, 8819-8830.
37. Murray, C. L. and Rabinowitz, J. C. (1981) *J. Biol. Chem.* 256, 5133-5161.