Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 1236–1240, March 1983
Biochemistry

Structure of a *Bacillus subtilis* bacteriophage SPO1 gene encoding a RNA polymerase σ factor

(regulatory protein/*middle* gene promoters/ribosome binding sites/gene cloning/DNA sequence determination)

MARIA COSTANZO AND JANICE PERO

The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138

Communicated by Jesse C. Rabinowitz, December 1, 1982

ABSTRACT **Gene 28 of *Bacillus subtilis* bacteriophage SPO1 codes for a regulatory protein, a σ factor known as σ^{gp28}, that binds to the bacterial core RNA polymerase to direct the recognition of phage *middle* gene promoters. *middle* promoters exhibit distinctive and conserved nucleotide sequences in two regions centered about 10 and 35 base pairs upstream from the start point of mRNA synthesis. Here we report the cloning of gene 28 and its complete nucleotide sequence. We infer that σ^{gp28} is a 25,707-dalton protein of 220 amino acids. Neither the nucleotide sequence of gene 28 nor the inferred amino acid sequence of σ^{gp28} exhibits extensive homology to the gene or protein sequence of *Escherichia coli* σ factor.**

The σ subunit of bacterial RNA polymerase (nucleosidetriphosphate:RNA nucleotidyltransferase, EC 2.7.7.6) plays a critical role in the initiation of transcription (1). The binding of σ to the core enzyme confers on polymerase the ability to recognize promoter sites on DNA which, in *Escherichia coli*, exhibit characteristic conserved sequences in two regions centered about 10 base pairs (bp) (the −10 region) and 35 bp (the −35 region) upstream from the start point of transcription. The consensus sequences on the nontranscribed strand for the −10 and −35 regions are T-A-T-A-A-T and T-T-G-A-C-A, respectively (2, 3).

In *E. coli*, a single 70,263-dalton σ factor is thought to initiate all transcription (4). In *Bacillus subtilis* and its bacteriophages, in contrast, multiple species of σ factor confer upon core RNA polymerase the ability to bind to different classes of promoters (5). RNA polymerase containing σ^{55}, the principal σ factor used by *B. subtilis* during vegetative growth, recognizes promoters that are similar in sequence to *E. coli* promoters (6, 7). During the lytic growth of certain bacteriophages and during the process of endospore formation, σ^{55} is replaced on the core RNA polymerase by novel phage and bacterial species of σ. The prototype for studies of novel σ factors is the product of bacteriophage SPO1 gene 28, a regulatory protein known as σ^{gp28}, which directs the recognition of phage *middle* gene promoters at intermediate times during the SPO1 lytic cycle (8–10).

Like *E. coli* promoters or promoters recognized by the principal form of *B. subtilis* RNA polymerase, *middle* gene promoters strongly conform to each other in two sequences centered at positions −35 and −10. The conserved −35 and −10 regions of *middle* gene promoters (A-G-G-A-G-A and T-T-T-N-T-T-T, respectively; T represents hydroxymethyluracil in SPO1 DNA) differ significantly, however, from the corresponding canonical hexamers of *E. coli* promoters or *B. subtilis* σ^{55}-controlled promoters (11, 12). Interestingly, this pattern of distinctive and conserved −10 and −35 regions is emerging as a general feature of promoters controlled by the various bacterial and phage-coded σs in *B. subtilis* (13–15). This argues that RNA polymerases with different species of σ factor see promoters in fundamentally the same manner but that different σ factors interact with different bases in both the −35 and −10 regions (5, 12).

How then do the structures of these novel σs compare with each other or with the structure of *E. coli* σ? The primary structure of *E. coli* σ has been derived from the nucleotide sequence of its cloned gene *rpo*D (4). Until now, the primary structure of any σ factor acting in *B. subtilis* has been unknown. Here we report the cloning of SPO1 gene 28, its complete nucleotide sequence, and the derived primary structure of σ^{gp28}.

MATERIALS AND METHODS

Construction of Recombinant Plasmids. SPO1 phages were grown and purified as described (8); gene 28 mutants *sus*F21 and *sus*33 were provided by S. Okubo (16). Phage and plasmid DNAs were purified and cleaved with restriction enzymes, and their fragments were isolated as described (17, 18) except that DNA fragments were eluted from agarose gels by the method of Vogelstein and Gillespie (19).

*Sau*3A-cleaved *Eco*RI* fragment 3 (2 μg) isolated from either wild-type SPO1 DNA (to construct pMC103) or gene 28 mutant *sus*F21 DNA (to construct pMC2) was cloned into *Bam*HI-cleaved pBR322 (0.2 μg). Conditions for *in vitro* ligation, transformation, and preparation of colony hybridization filters were as described (20) except that T4 DNA ligase was purchased from Bethesda Research Laboratories and *E. coli* strain HB101 was the transformation recipient. DNA for use as a hybridization probe (see *Results*) was radioactively labeled by nick-translation (21). Hybridizations were performed as described (18) except that a 200-fold excess of unlabeled pBR322 DNA was added to the hybridization mixture for pMC103 screening.

***In Vitro* Transcription-Translation.** Coupled *in vitro* transcription-translation was performed with an S150 extract, ribosomes, and initiation factors purified from *E. coli* MRE 600 cells according to the method of Schweiger and Herrlich (22). DNA (1–2 μg) was used in a 50-μl reaction mixture containing 15 μCi (1 Ci = 3.7×10^{10} Bq) of [^{35}S]methionine (from New England Nuclear). Samples were precipitated with trichloroacetic acid, resuspended in sample buffer, and run on 12% polyacrylamide gels by the method of Laemmli (23).

DNA Sequence Determination. Restriction fragments were radioactively labeled at their 5′ ends by the phosphate exchange kinase reaction with T4 polynucleotide kinase (from Boehringer Mannheim) (24) or at their 3′ ends by extension with the Klenow fragment of *E. coli* DNA polymerase I (from Bethesda Research Laboratories) (25). Nucleotide sequence was determined as described by Maxam and Gilbert (26).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s).

Purification and NH_2-Terminal Amino Acid Sequence of gp28. A plasmid capable of overproducing gp28 in *E. coli* was constructed by ligating the *Sau*3A insert from pMC103 into the *Bam*HI site of pλ8, a pBR322 derivative containing a 1.1-kilobase (kb) fragment of bacteriophage λ DNA bearing the strong promoter P_L and the *N* gene (27). Induction of transcription from P_L, by inactivation of the temperature-sensitive λ repressor carried by the host strain, resulted in the production of gp28 as a significant fraction of total cell protein. gp28 was purified to apparent homogeneity from extracts of induced cells by DEAE-Sephacel and Sephacryl S-200 chromatography. The sequence of the first five NH_2-terminal amino acids of gp28 was determined on a Beckman sequencer with approximately 50 μg of gp28. The overproduction, purification, and NH_2-terminal amino acid sequence of gp28 will be described in detail elsewhere.

RESULTS

Cloning of Mutant Gene 28. Because σ factors that interact with *B. subtilis* core RNA polymerase can also alter the promoter recognition specificity of *E. coli* RNA polymerase (28, 29), we considered the possibility that a plasmid carrying a copy of SPO1 gene 28 would be lethal to its *E. coli* host. To avoid this potential problem, we first cloned a mutant version of gene 28 (although, as shown below, this precaution proved to be unnecessary). Coupled transcription and translation experiments with isolated SPO1 DNA fragments by Chelm *et al.* (30) and our own similar experiments (see Fig. 2*B*, lane 2) mapped gene 28 to a 2-kb *Bgl* II–*Bst*EII restriction fragment contained within *Eco*RI* fragment 3 (Fig. 1). To clone DNA from this region of the genome, *Eco*RI* fragment 3 was isolated from gene 28 mutant *sus*F21 phage DNA, digested with *Sau*3A, and ligated into the *Bam*HI site of pBR322. Ampicillin-resistant tetracycline-sensitive transformants were selected and screened by *in situ* colony hybridization. The probe used for the colony hybridization was a 3.4-kb *Bgl* II–*Bst*NI fragment isolated from *Eco*RI* fragment 3 and radioactively labeled by nick-translation; this larger gene 28-containing segment could be obtained in pure form more readily than the smaller *Bgl* II–*Bst*EII fragment (Fig. 1).

One colony to which the probe hybridized contained a plasmid, pMC2, with two inserts: *Sau*3A fragments of 1.2 kb and 140 bp. To verify that these inserts were indeed derived from the 3.4-kb *Bgl* II–*Bst*NI region, isolated pMC2 DNA was radioactively labeled by nick-translation and hybridized to a Southern blot of *Eco*RI* fragment 3 cleaved with both *Bgl* II and *Bst*NI. pMC2 DNA hybridized strongly to the 3.4-kb *Bgl* II–*Bst*NI fragment (not shown); no significant hybridization to any other DNA fragment was observed.

We examined the proteins produced from the pMC2 inserts in an *E. coli* coupled *in vitro* transcription-translation system. To eliminate a background of pBR322-encoded proteins, the plasmid was first digested with *Hae* III—which cuts most DNAs, including pBR322, into small pieces, but rarely cuts SPO1 DNA [only 5 sites exist in the 140-kb genome (31)]. *Hae* III digestion left the pMC2 inserts intact and flanked by short (25 and 79 bp) segments of pBR322. Lane 5 of Fig. 2*B* shows that no proteins were produced from *Hae* III-digested pBR322. Lane 4 shows that the major product of *Hae* III-digested pMC2 was a protein of approximately 21,000 daltons. This corresponds in size to the nonsense fragment of gp28 produced when *Eco*RI* fragment 3 from the gene 28 mutant *sus*F21 was subjected to *in vitro* transcription-translation (Fig. 2*A*). *Sau*3A digestion did not affect the appearance of this protein (not shown), indicating that its coding sequences must be entirely contained within the 1.2-kb *Sau*3A insert.

Cloning of Wild-Type Gene 28. The evidence that the pMC2 insert contained mutant gene 28 seemed sufficiently encouraging to attempt to repeat the cloning experiment but starting with *Eco*RI* fragment 3 from wild-type SPO1 DNA. The radioactive probe used to screen the ampicillin-resistant tetracycline-sensitive transformants for clones of wild-type gene 28 was the 1.2-kb *Sau*3A insert isolated from pMC2 DNA. One of the colonies to which the probe hybridized contained a plasmid, pMC103, with a single 1.2-kb *Sau*3A insert whose restriction map was found to be identical to that of pMC2 (Fig. 3), proving that we had in fact cloned the identical fragment from wild-type DNA. If this fragment did contain gene 28, we expected that the 21,000-dalton protein produced from pMC2 in the *in vitro* transcription-translation system would be replaced in pMC103-directed protein synthesis with the full-size gene 28 product (26,000 daltons). This expectation was confirmed in the experiment (lane 3 of Fig. 2*B*), establishing that pMC103 did contain gene 28.

Marker Rescue of Gene 28 Mutants by pMC103 and pMC2. Genetic evidence that pMC103 and pMC2 contained SPO1 gene 28 was provided by marker rescue experiments. Competent cells of a nonpermissive host strain were transformed with linear plasmid DNA and, shortly thereafter, were infected with nonsense mutant phage. If the plasmid DNA contained the wild-

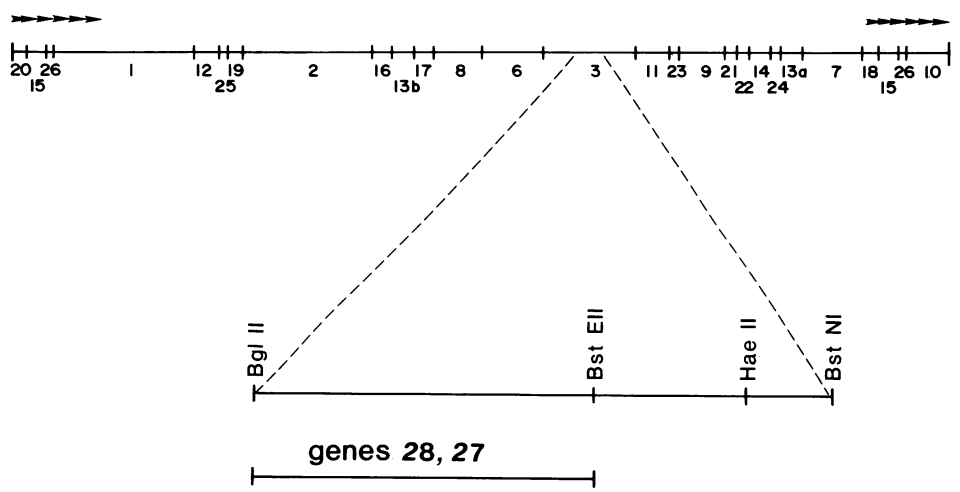

FIG. 1. *Eco*RI* restriction map of SPO1 DNA (31). The genome is approximately 140 kb with 12.4-kb direct terminal repeats. Below is shown the gene 28 region, as mapped by Chelm *et al.* (30), except that gene 28 is located to the left of gene 27 (see *Discussion*).

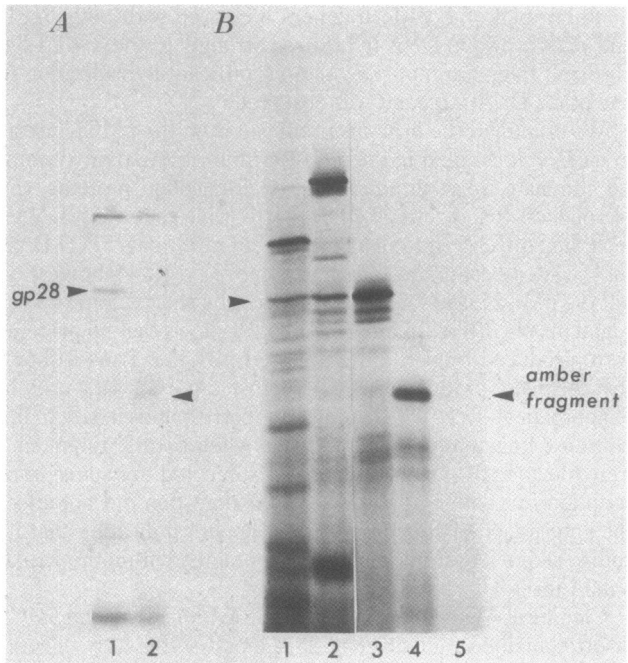

FIG. 2. Proteins produced from *in vitro* transcription-translation of recombinant plasmids and of SPO1 DNA. Proteins synthesized from the indicated DNAs (1–2 μg) were radioactively labeled with [^{35}S]methionine (15 μCi) and separated by gel electrophoresis. (*A*) Identification of gp*28*. Lanes: 1, *Eco*RI* fragment 3 from wild-type SPO1 DNA; 2, *Eco*RI* fragment 3 from the DNA of the gene *28* nonsense mutant *sus*F21. Arrowheads indicate the positions of gp28 and the *amber* fragment. The band identified as gp28 also comigrated with unlabeled gp28 purified from infected cells (not shown). (*B*) Proteins produced from wild-type and mutant gene *28* clones. Lanes: 1, isolated *Eco*RI* fragment 3; 2, isolated 2-kb *Bgl* II–*Bst*EII subfragment of *Eco*RI* fragment 3; 3, *Hae* III-digested pMC103; 4, *Hae* III-digested pMC2; 5, *Hae* III-digested pBR322.

type allele of the phage mutation, recombination between the two could produce a wild-type genome, thus allowing plaques to form on a lawn of the nonpermissive host.

Table 1 shows the results of marker rescue experiments using pMC103 and pMC2 DNA to rescue the infections of two different gene *28* nonsense mutants (*sus*F21 and *sus*33). In no case did transformation with pBR322 DNA alone increase the background number of plaques per plate, whereas pMC103 rescued the infections of both mutant phages, *sus*F21 and *sus*33, as efficiently as did wild-type SPO1 DNA. Even pMC2, constructed from the DNA of *sus*F21 phage, gave a small but significant increase in the number of plaques per plate when used to rescue the infection of *sus*33 phage. It was expected that the increase would be small because rescue of one gene *28* allele by DNA of a different allele requires that recombination take place within the gene. Thus, these experiments proved definitively that pMC103 and pMC2 contain SPO1 gene *28*.

Nucleotide Sequence of Gene 28. The nucleotide sequence of gene *28* and its 5′ and 3′ flanking regions was determined by the Maxam–Gilbert chemical method (26). Restriction fragments were radioactively labeled either at their 5′ ends by T4 polynucleotide kinase or at their 3′ ends by the Klenow fragment of T4 DNA polymerase. The strategy is presented in Fig. 3, and the DNA sequence of the gene and the derived protein sequence of gp28 are shown in Fig. 4. DNAs from both pMC2 and pMC103 were used for sequence analysis, in some cases as a mixture; the inserts were found to be identical at all positions except at nucleotide 851 where C in pMC103 is replaced by T in pMC2, changing a CAG codon for glutamine into the *amber* stop codon TAG.

Location of the Translational Initiation Codon for gp28. The cloned insert in pMC103 has only a single open reading frame in either direction large enough to encode a polypeptide the size of gp28, 26,000 daltons. Furthermore, the position of the *amber* mutation in gene *28* establishes this to be the correct reading frame. Near the 5′ end of this 243-codon open reading frame there is a translational initiation codon, ATG, preceded at the correct distance by a strong ribosome binding site [$\Delta G = -14.6$ kcal/mol (1 cal = 4.184 J) as calculated according to the rules of Tinoco *et al.* (34)]. In accordance with the hypothesis of McLaughlin *et al.* (35), the ribosome binding site has extensive complementarity to the 3′ end of *B. subtilis* 16S ribosomal RNA, making the interaction between the mRNA and rRNA stronger than that usually found in *E. coli*.

To confirm that this site is actually the translational start point of gp28, the sequence of the NH_2-terminal five amino acids of gp28 was determined. The sequence obtained, Met-Val-Glu-Asn-Val, is in perfect agreement with the NH_2-terminal amino acid sequence predicted from the nucleotide sequence of gene *28*. This pattern of five amino acids is not found anywhere else in the reading frame of gp28.

A *middle* Gene Promoter Located After Gene 28 and Before Gene 27. *middle* gene promoters have highly conserved sequences at both the −35 and −10 regions (A-G-G-A-G-A and T-T-T-N-T-T-T, respectively) (11, 12). Inspection of the pMC103 nucleotide sequence for regions homologous to the *middle* gene promoter consensus sequence revealed a perfect match immediately downstream from the 3′ end of gene *28* (Fig. 4). Transcription of this region of SPO1 DNA, cut with various restriction enzymes, by σ^{gp28}-containing RNA polymerase produced

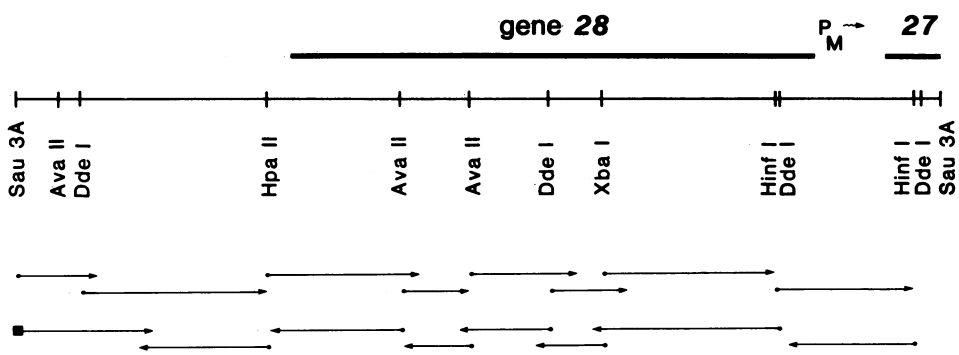

FIG. 3. Restriction map and sequence-determination strategy for pMC103 and pMC2. The location of gene *28* and the tentative location of gene *27* are indicated; the direction of transcription through this region is from left to right. Arrows indicate direction and extent of sequence analysis; arrows ending with solid circles denote sequence analysis of 5′ end-labeled fragments; arrow with a square denotes sequence analysis of a 3′ end-labeled fragment.

Table 1. Marker rescue of gene *28* mutants by pMC103 and pMC2

DNA	Plaques, no./plate	
	*sus*F21	*sus*33
—	7	13
pBR322	6	14
pMC103	132	316
pMC2	6	33
SPO1†	141	429

The values listed represent means for the two mutant phages. Marker rescue assays were performed as described by Cregg and Stewart (32) except that competent cells were prepared by the method of Dubnau and Davidoff-Abelson (33).

† SPO1 DNA was digested with *Eco*RI* and failed to transfect.

runoff RNAs of the sizes predicted if this sequence were a functional *middle* promoter (data not shown). Downstream from the *middle* promoter is a strong ($\Delta G = -16.8$ kcal/mol) ribosome binding site followed by a translational open reading frame containing the initiation site TTG [UUG and GUG as well as AUG function as initiation codons in *B. subtilis* (35)]. This is most likely the 5′ end of gene *27* because our marker rescue experiments with overlapping segments of DNA cloned from this region (unpublished data) have established that gene *27* is located to the right of gene *28* and that its 5′ end is located on pMC103 close to the 3′ end of gene *28*.

DISCUSSION

We have determined the complete nucleotide sequence of SPO1 gene *28* and its 5′ and 3′ flanking regions. NH_2-terminal amino acid sequence analysis of purified gp28 was used to locate the 5′ end of gene *28*. σ^{gp28} is composed of 220 amino acid residues and has a molecular mass of 25,707 daltons, which is close to the value, 26,000 daltons, determined in this laboratory by sizing on polyacrylamide gels (36). Thus, unlike *E. coli* σ factor whose electrophoretic mobility is anomalously low (4), σ^{gp28} migrates at a rate consistent with its actual size.

σ^{gp28} is a highly charged protein, with 34% charged residues as compared to 24% for an "average" protein [this value was derived from a comparison of the amino acid compositions of 314 proteins (37)]. In this respect it is similar to *E. coli* σ which has 35% charged residues (4). However, *E. coli* σ has more acidic than basic residues (20.4% vs. 14.5%) and the acidic residues are concentrated in the NH_2-terminal third of the protein,

```
                                                                                                    100
GATCACAGACAAAGAAGGTAAGGTAGTATTAGAAGTTATCAAAGAGGTAGGGGTCCTTGACACATCTGCAACACAGGAGCTTAGAGTAAACATTGTAAAA
                                                                                                    200
GACGACAAAGGAAGAGTAAGTAGCTATTCAAAAGTTCTGGAGAAAACATAAGGATGAAGAGTGGAAGGTTGGTAAAGGTCTATTTGTAACTGGAAGAGAC
                                                                                                    300
GCAGTAAATTTAAGCCCTTTATTGGAACAAGCGGGCAACGAGATTATACACATTAGATAAGACAAGCAACATGGGGGAACTGCGGTTGGAAAGAGACAAA

ATTAAAGCCGGAGGAGAAAAAGAAA ATG GTA GAA AAT GTA ACG TAT TCA GAG GAT GGT CAA TTA CTG AAA CTG AAA TGG
                          met val glu asn val thr tyr ser glu asp gly gln leu leu lys leu lys trp
                        400
GAA GAT GTG TAC GAA CAA TTT AAA AAT CTC ATC ACA TTT GCG GCT AGA CAG CAG ATG GAG AAC AAT GGA GCA GAC
glu asp val tyr glu gln phe lys asn leu ile thr phe ala ala arg gln gln met glu asn asn gly ala asp
                                                            500
ACC ATG ATG TCT AGG CAG GAC CTA GAG CAG GAG GGT CTA CTA AAA CTG TAT GAC TGC TGG GAA AAA TGG TGC TTC
thr met met ser arg gln asp leu glu gln glu gly leu leu lys leu tyr asp cys trp glu lys trp cys phe
                                                                                        600
AAA GAG AAT AAA CAG ATG GAT GAG TTC GGT CCC ATC TTC CGT AAA TCC CTA TTT AGA AAA GTA AAA CAA AGT GGG
lys glu asn lys gln met asp glu phe gly pro ile phe arg lys ser leu phe arg lys val lys gln ser gly

GGA ACT GGT AGA GCA CTT GGC TTT GTG GCT ATT GAT GAT GAG GAC AAC CCC TTA GAG AAC ATG CTC AAA GAC GAA
gly thr gly arg ala leu gly phe val ala ile asp asp glu asp asn pro leu glu asn met leu lys asp glu
                        700
AAC ACA GTG GAT GTT GTA GAG AAA ATA CAC TTT TCA GAG GGT CTA GAA AAG CTC AAA GAG ACA TTA GAA AGT GAT
asn thr val asp val val glu lys ile his phe ser glu gly leu glu lys leu lys glu thr leu glu ser asp
                                                            800
ATC GCA AAA TCA CTG CTA GAG GAA CTT ATC AAC CCT AGT GAC CAG ACC ATT TAT AAC GTG TTG ATT GAT ATT GAG
ile ala lys ser leu leu glu glu leu ile asn pro ser asp gln thr ile tyr asn val leu ile asp ile glu
                                                                                        900
CGT AAA AAG ATG CTT AAG TCA CAG GGT CAT AGG GTT AAT GTG CCA AAA GAT ACA ACA GTA CGC ATG AAG CAT ATT
arg lys lys met leu lys ser gln gly his arg val asn val pro lys asp thr thr val arg met lys his ile

GAC CAA ACA CTA GGC ATA TCC AAC AAG CAA TAT GAC TCA GAA TTG AAG AAA TTC GTG AAG CGG CTC ACT ATC TAC
asp gln thr leu gly ile ser asn lys gln tyr asp ser glu leu lys lys phe val lys arg leu thr ile tyr
                    1000
ACT ATC TAG GAGTACTAGGATAAGTACAGGAGAAAAAACATAAACAGAGAATTTGTTTTATTTGAGAGAATTAGGAAAAGGGAGAGTACAA TTG
thr ile ---                                                                                      met
                                    1100
TCA ATC ACA GCC ATG GAT GCC AAG TTA CAA
ser ile thr ala met asp ala lys leu gln
```

FIG. 4. Nucleotide sequence of gene *28* and surrounding DNA. Only the nontranscribed DNA strand is shown. Beneath the DNA sequence is the predicted amino acid sequence of gp28. The ribosome binding site for gp28 and the putative ribosome binding site for gp27 are underlined. The −35 and −10 regions of the intercistronic *middle* promoter are enclosed in dotted lines.

whereas σ^{gp28} has approximately equal percentages of acidic and basic residues (17.7% and 15%, respectively) and they are evenly distributed throughout the protein.

The secondary structure of σ^{gp28} was predicted by the method of Chou and Fasman (38). The results of this analysis suggest that roughly 70% of the protein is in the α-helix conformation, with more than half the protein in four long helices of about 30 amino acids each. Approximately 16% of σ^{gp28} is predicted to be β-pleated sheet and about 9%, reverse turn. *E. coli* σ is also predicted to have a high content of α-helix (50–55%) (4).

A comparison, by computer (39), of the nucleotide sequences of gene 28 and the *E. coli* σ factor gene *rpo*D and of their derived amino acid sequences failed to reveal any extensive homology between the two genes or proteins. However, several short regions of amino acid homology were observed, the best examples being two stretches of five identical amino acids out of six. The significance of these regions of homology is unclear. Perhaps they reflect common structural features of σ factors that are required for interaction with DNA or with core RNA polymerase. In fact, both σ factors can bind to the same core RNA polymerases: *E. coli* σ can interact with *B. subtilis* core enzyme (28), and the 28,000-dalton protein of phage SP82 (the σ^{gp28} analog of this closely related phage) can direct *E. coli* RNA polymerase to bind *middle* gene promoters (29).

How might the expression of gene 28 itself be controlled? Gene 28 is located downstream from a weak *early* promoter; the 2.7-kb transcript produced from this promoter presumably encodes σ^{gp28} (30). This mRNA must also encode gp27 because we have located gene 27 immediately downstream from gene 28 (Fig. 3). Interestingly, we have located a *middle* promoter in the region between these two genes. This suggests that gene 27, whose product is involved in both DNA replication and late transcription (40), is transcribed both at early and middle times after infection under the control of two different promoters. Chelm *et al.* (30) previously identified a *middle* gene promoter in or about the same position as the one identified here by an electron microscopic analysis of gp28-containing RNA polymerase bound to DNA fragments. Because, until now, gene 27 had been thought to precede gene 28, they suggested that gene 28 might be transcribed from this *middle* promoter, thus partially regulating its own expression. In light of the revised gene order, the question of whether gp28 regulates transcription of its own gene remains unanswered.

Above we raised the possibility that a plasmid carrying a copy of gene 28 might be lethal for *E. coli*. Our present experiments leave this question unresolved because we consider it unlikely that gene 28 is expressed from pMC103. The orientation of gene 28 in pMC103 is opposite to that of the promoter for the *tet* gene of pBR322 and in preliminary experiments we were unable to detect an early promoter for gene 28 on the insert.

In summary, we have determined the primary structure of a σ factor that interacts with *B. subtilis* core RNA polymerase. To date six different phage or bacterial σs that confer unique promoter recognition properties have been described (reviewed in ref. 5). It will be interesting to compare the structure of σ^{gp28} with that of the other σs to see whether they are related evolutionarily.

We gratefully thank Nancy Hannett for providing certain DNA fragments and SPO1 DNA, Klaus Koller and L. Bogorad for help in setting up the *in vitro* transcription-translation system, Roger Yocum for advice on protein sequence determination, Jim Pustell and Carol Stoner for help with the computer programs, and Richard Losick for help in the preparation of this manuscript. M.C. was a Predoctoral Fellow of the National Science Foundation. This work was supported by National Science Foundation Grant PCM 7821570 to J.P.

1. Chamberlin, M. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 17–67.
2. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
3. Siebenlist, U., Simpson, R. B. & Gilbert, W. (1980) *Cell* **20**, 269–281.
4. Burton, Z., Burgess, R. R., Lin, J., Moore, D., Holder, S. & Gross, C. A. (1981) *Nucleic Acids Res.* **9**, 2889–2903.
5. Losick, R. & Pero, J. (1981) *Cell* **25**, 582–584.
6. Lee, G., Talkington, C. & Pero, J. (1980) *Mol. Gen. Genet.* **180**, 57–65.
7. Moran, C. P., Jr., Lang, N., LeGrice, S. F. J., Lee, G., Stephens, M., Sonenshein, A. L., Pero, J. & Losick, R. (1982) *Mol. Gen. Genet.* **186**, 339–346.
8. Fox, T. D., Losick, R. & Pero, J. (1976) *J. Mol. Biol.* **101**, 427–433.
9. Duffy, J. J. & Geiduschek, E. P. (1977) *Nature (London)* **270**, 28–32.
10. Talkington, C. & Pero, J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1185–1189.
11. Talkington, C. & Pero, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5465–5469.
12. Lee, G. & Pero, J. (1981) *J. Mol. Biol.* **152**, 247–265.
13. Moran, C. P., Jr., Lang, N. & Losick, R. (1981) *Nucleic Acids Res.* **9**, 5979–5990.
14. Gilman, M. Z., Wiggs, J. L. & Chamberlin, M. J. (1981) *Nucleic Acids Res.* **9**, 5991–6000.
15. Unnasch, N. (1982) Dissertation (Harvard Univ., Cambridge, MA).
16. Okubo, S., Yanagida, T., Fujita, D. J. & Ohlsson-Wilhelm, B. M. (1972) *Biken. J.* **15**, 81–97.
17. Talkington, C. & Pero, J. (1977) *Virology* **83**, 365–379.
18. Lee, G., Hannett, N. M., Korman, A. & Pero, J. (1980) *J. Mol. Biol.* **139**, 407–422.
19. Vogelstein, B. & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 615–619.
20. Segall, J. & Losick, R. (1977) *Cell* **11**, 751–761.
21. Rigby, P. W. J., Duckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 239–251.
22. Schweiger, M. & Herrlich, P. (1974) *Curr. Top. Microbiol. Immunol.* **65**, 59–132.
23. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
24. Berkner, K. L. & Folk, W. R. (1977) *J. Biol. Chem.* **252**, 3176–3184.
25. Klenow, H. & Henningson, I. (1970) *Proc. Natl. Acad. Sci. USA* **65**, 168–175.
26. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–559.
27. Reed, R. R. (1981) *Cell* **25**, 713–719.
28. Shorenstein, R. G. & Losick, R. (1973) *J. Mol. Biol.* **248**, 6170–6173.
29. Achberger, E. C. & Whiteley, H. R. (1980) *J. Biol. Chem.* **225**, 11957–11964.
30. Chelm, B. K., Romeo, J. M., Brennan, S. M. & Geiduschek, E. P. (1981) *Virology* **112**, 572–588.
31. Pero, J., Hannett, N. M. & Talkington, C. (1979) *J. Virol.* **31**, 156–171.
32. Cregg, J. M. & Stewart, C. R. (1978) *Virology* **86**, 530–541.
33. Dubnau, D. & Davidoff-Abelson, R. (1971) *J. Mol. Biol.* **56**, 209–221.
34. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. J. (1973) *Nature (London) New Biol.* **46**, 40–41.
35. McLaughlin, J. R., Murray, C. L. & Rabinowitz, J. C. (1981) *J. Biol. Chem.* **256**, 11283–11291.
36. Pero, J., Nelson, J. & Fox, T. D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1589–1593.
37. Dayhoff, M. O., Hunt, L. T. & Hurst-Calderone, S. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Found., Washington, DC), Suppl. 3, Vol. 5, pp. 363–369.
38. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–148.
39. Pustell, J. & Kafatos, F. C. (1982) *Nucleic Acids Res.* **10**, 4765–4782.
40. Greene, J. R., Chelm, B. K. & Geiduschek, E. P. (1982) *J. Virol.* **41**, 715–720.