Nucleotide Sequence of the Temperate *Bacillus subtilis*
Bacteriophage SPO2 DNA Polymerase Gene L

BENNY RÁDÉN AND LARS RUTBERG*

Department of Bacteriology, Karolinska Institutet, S-104 01 Stockholm, Sweden

Received 15 February 1984/Accepted 6 May 1984

Temperate *Bacillus subtilis* phage SPO2 codes for a phage-specific DNA polymerase. The polymerase gene has been cloned, and its nucleotide sequence has been determined. Within the sequence there is an open reading frame starting with a TTG and ending with three consecutive translational stop codons. Ten base pairs upstream from the proposed TTG initiation codon there is a probable ribosome-binding site with a calculated free energy of interaction with the 3’ end of *B. subtilis* 16S rRNA of -15 kcal (-63 kJ/mol). Based on the sequence and the expression of the polymerase gene in three different hybrid plasmids, we conclude that this open reading frame is the structural gene for SPO2 DNA polymerase. The predicted molecular weight of the polymerase is 72,486. In hybrid plasmid pJB74, the terminal triplet of an open reading frame with coding capacity for a protein of ca. 10 kilodaltons overlaps with the translational initiation triplet TTG of the gene. We speculate that transcription and translation of this open reading frame can influence the amount of phage DNA polymerase made in SPO2-infected bacteria.

SPO2 is a temperate *Bacillus subtilis* bacteriophage (3). Some 20 genes have been identified in SPO2 by complementation and ordered on a linear genetic map (38). SPO2 and most virulent *B. subtilis* phages grow in the presence of 6-(p-hydroxyphenylazo)-uracil, an inhibitor of *B. subtilis* DNA polymerase III (4). It is proposed that 6-(p-hydroxyphenylazo)-uracil-resistant phages produce their own DNA polymerase and replicate independently of DNA polymerase III (12). A new DNA polymerase has been found in SPO2-infected bacteria; it is a product of a complementation group called gene *L* (28).

Recently we cloned the wild-type allele of gene *L* mutation *susL244* in plasmid pC194 (30). Three hybrid plasmids were isolated (originally called pC194-74, pC194-91, and pC194-96 [30]); one of these, pJB96, causes a 10- to 20-fold increase in the level of in vitro DNA polymerase activity in a *B. subtilis* *polA(Ts)* *polC(Ts)* mutant, indicating that it contains a complete gene *L*. For the other two plasmids a smaller increase (pJB74) or no increase (pJB91) in DNA polymerase activity was found. Gene *L* from pJB96 has been subcloned in *Escherichia coli* (31), where it is also expressed, which demonstrates that the gene codes for a new DNA polymerase rather than an activator which modifies a *B. subtilis* enzyme.

Restriction site maps of pJB74, pJB91, and pJB96 show that all three plasmids contain a complete gene *L* (this paper). To understand the different levels of expression of gene *L* in the hybrid plasmids, we determined its nucleotide sequence. The predicted molecular size of the enzyme is 72 kilodaltons (kd). The results indicate that the polymerase is translated from a polycistronic messenger and that its synthesis may be controlled both at the transcriptional and translational level.

**MATERIALS AND METHODS**

**Bacteria and phage.** The following *B. subtilis* strains were used: 3G18 (ade met trpC2), BD294 (trpC2 thr-5 polA59 polC59), SR135 (trp-7 spoOA9 sup-3), and W168 (prototrophic). Strain BD294 was obtained from T. A. Trautner.

Phage SPO2C1 was originally obtained from J. Marmur.

SPO2 *susL244* (38) was obtained from T. A. Trautner. The bacteria were kept on TBAP plates (trypic blood agar base; Difco Laboratories). For strains carrying pC194 or pUB110, chloramphenicol (20 µg/ml) or kanamycin (5 µg/ml), respectively, was added to the plates. SPO2 was assayed on NY plates as described (28). Liquid cultures were grown in NY broth with added antibiotic when appropriate. Phage stocks were prepared in liquid culture or on plates as described (28). *Escherichia coli* JM103 [Δlac pro]? thiE stra supE6 endA sbcB hsdR<sup>F</sup> traD36 proAB lac[:F<sup>+</sup>] ΔM15] and M13mp8 and M13mp9 were obtained from L.-O. Hedén. The methods for growing M13 and preparing RF forms were essentially as described by Messing et al. (19-21).

**Plasmids and preparation of plasmid DNA.** pC194 is a 2.9-kilobase-pair *Staphylococcus aureus* plasmid (14, 16) which can be maintained in *B. subtilis* (8); it confers chloramphenicol *Cm* resistance. The isolation of pC194-SPO2 hybrids has been described (30). pUB110 is a 4.5-kilobase-pair plasmid originally described by Gryczan et al. (10, 11); it confers kanamycin *Km* resistance. Plasmid DNA was prepared as described by Canosi et al. (5), except that preparation of a clear lysate was omitted and instead the whole lysate was used in the first cesium chloride-ethidium bromide centrifugation. In our hands this considerably improves the recovery of covalently closed circular plasmid DNA. Minipreparations of plasmid were prepared as described by Birnboim and Doly (2). DNA concentrations were determined on agarose plates containing 1 µg of ethidium bromide per ml with calf thymus DNA as the standard (17).

**Transformation.** For transformation of *B. subtilis* with plasmid DNA, competent cells (1) or protoplasts (6) were used. *E. coli* transfection was done as described by Maniatis et al. (17).

**Electrophoresis.** DNA restriction fragments were analyzed by using 0.8 to 2% agarose gels or 5% polyacrylamide-bisacrylamide (29:1) gels. For sequencing, the samples were electrophoresed on 8 and 6% polyacrylamide-bisacrylamide (19:1) gels with 8 M urea. The buffer system used was 90 mM Tris base–90 mM boric acid–1 mM EDTA (pH 8.3).

**Structure of pJB74, pJB91, and pJB96.** SPO2 gene *L* was originally cloned from partially HindIII-digested phage DNA (30). Four HindIII fragments are common to all plasmids;
pJB91 contains one and pJB74 contains two additional fragments (Fig. 1). The order of the HindIII fragments was determined from single and double digestes with BglII, HincII, HindIII, and PvuII (Fig. 2). The order of the four common HindIII fragments is the same in the three plasmids, but their orientation relative the vector pC194 is inverted in pJB96 compared with pJB74 and pJB91. Unique EcoRI and PstI sites are located in pJB74 HindIII fragment E.

**Sequencing.** DNA sequencing was done by the dideoxy method (32) by using the M13 cloning system (19). Template single-stranded DNA was prepared by infecting exponentially growing JM103 with phage and incubating with shaking at 37°C for 12 to 18 h. The bacteria were then pelleted by centrifugation in an Eppendorf centrifuge, and M13 single-stranded DNA prepared from the supernatant as described by Messing (21).

The sequencing strategy was based on the structure of pJB74 (see Fig. 2). HindIII fragments E, B, F, and G were isolated from agarose or polyacrylamide gels, electroeluted, and then purified on a Whatman DE-52 column. The 2,950-base-pair (bp) PvuII fragment, the 1,950-bp HindII fragment, and the 445-bp PstI-PvuII fragment (which overlaps HindIII fragments E and B) were similarly isolated and purified. Fragments E, B, F, G, the 2,950-bp PvuII fragment, and the 1,950-bp HindII fragment were then cloned in both directions into M13mp8. The two EcoRI-HindIII subfragments (E' and E") of HindIII fragment E and the 445-bp PstI-PvuII fragment were cloned into M13mp9. With the above constructions, the areas indicated by dotted lines in Fig. 3 could be sequenced, but the order of fragments F and G could not be determined.

To sequence the rest of gene L, a modification of the method of Poncz et al. (26) was used. Thirty micrograms of the RF form of M13mp8 carrying the 2,950-bp PvuII fragment was linearized by cleaving with EcoRI. The DNA was

---

**FIG. 1.** Polyacrylamide gel electrophoresis pattern of HindIII-cleaved plasmid DNA. Lane 1, pJB74; lane 2, pJB91; lane 3, pJB96; lane 4, SPO2 DNA. EcoRI-cleaved φ105 DNA (33) and HaeIII-cleaved pBR322 (35) were used as size markers.

**FIG. 2.** Restriction maps of pC194-SPO2 hybrid plasmids. The double line represents pC194, and the single line represents SPO2 DNA. Symbols used for restriction enzyme sites are: B, BglII; E, EcoRI; H, HincII; H, HindIII; P, PstI; P, PvuII; and S, Sau3A. The PvuII site in pC194 is actually two sites separated by 40 bp as determined by restriction mapping and sequencing (L. Melin, unpublished data). The Sau3A mapping is not complete, but the two sites relevant for the construction of pTB9 are shown (see Fig. 4). The relative order of HindIII fragments F and G was determined by nucleotide sequencing. The capital letters below the lines refer to the SPO2 HindIII fragments. CAT, Chloramphenicol transacylase; the arrow indicates the transcriptional direction (13).
extracted with phenol, ethanol precipitated, and then suspended in 20 mM Tris-hydrochloride (pH 8.0), 12 mM CaCl₂, 12 mM MgCl₂, 600 mM NaCl, and 1 mM EDTA. Six units of nuclease Bal 31 was added to the sample at 37°C. At 1-min intervals, 2-μl samples were withdrawn into 1 μl of 0.5 M EDTA on ice to stop the reaction. The extent of the reaction was monitored by electrophoresing aliquots on 1% agarose gels. The samples were pooled, extracted with phenol, ethanol precipitated, and suspended in 50 μl of 10 mM Tris-hydrochloride (pH 7.5)-10 mM MgSO₄-1 mM dithiotreitol-50 mM NaCl-1 mM each of the four deoxyribonucleotides (dATP, dCTP, dGTP, dTTP). One unit of the Klenow fragment of *E. coli* DNA polymerase I was added, and the sample was incubated at 25°C for 15 min. The sample was then heated to 70°C for 5 min to destroy the DNA polymerase, and after being cooled to 37°C, it was cleaved with *PstI*. The DNA was electrophoresed on a vertical 0.8% agarose gel. Fragments smaller than 2,950 bp were cut out of the gel and sliced into eight consecutive fractions. DNA was electroeluted from each fraction, purified, and cloned into M13mp9 cleaved with *PstI* and *HincII*. The DNA was used to transfect JM103, and phage carrying inserts was identified on plates containing isopropyl-β-d-thiogalactopyranoside and 5-mono-4-chloro-3-indolyl-β-d-galactopyranoside (22). In this manner we could generate a series of clones carrying progressively shortened fragments of the 2,950-bp *PvuII* fragment in a fixed orientation. A preliminary characterization of the clones could be done by hybridizing DNA from the clones with DNA from some of the previously cloned fragments, e.g., *HindIII*-G or the 1,950-bp *HincII* fragment. Sequence analysis of clones from the eight groups gave overlapping segments (10 to 99% overlap) throughout the whole of the 2,950-bp *PvuII* fragment. A total of 10 Bal 31-derived fragments were sequenced, covering a region extending 781 bp into the A fragment. From the overlaps the relative order of *HindIII* fragments G and F could be determined.

**Enzymes.** Restriction enzymes were obtained from New England Biolabs or Boehringer Mannheim Biochemicals. The reaction conditions were as described by Maniatis et al. (17), T4 ligase was obtained from New England Biolabs, and the Klenow fragment of *E. coli* DNA polymerase I was obtained from Boehringer. SPO2 DNA polymerase was assayed as previously described (30).

**Chemicals.** Calf thymus DNA, ethidium bromide, bovine pancreas DNase, lysozyme, isopropyl-β-d-thiogalactopyranoside, 5-mono-4-chloro-3-indolyl-β-d-galactopyranoside, and 2-mercaptoethanol were obtained from Sigma Chemical Co. Unlabeled deoxyribonucleotides and deoxyribonucleotides were obtained from Sigma or P-L Biochemicals. [*α-32P]*dCTP was from New England Nuclear Corp.; [*α-32P]*dGTP was from New England Nuclear or Amersham Corp. Agarose, polyacrylamide, and bisacrylamide were from Bio-Rad Laboratories. Polyethylene glycol 6000 was from BDH. CsCl was from BDH or Merck & Co., Inc. All

---

**FIG. 3.** Strategy for sequencing gene *L* from pJB74. The upper line represents the SPO2 part of pJB74. Above the line are shown the restriction enzymes sites (symbols are as described in the legend to Fig. 2). The capital letters below the line denote the SPO2 *HindIII* fragments. The thin continuous lines represent restriction fragments cloned in M13mp8 and M13mp9. The dotted lines and the arrows denote the parts of the fragments that were sequenced and the direction of sequencing. The dotted line marked Bal 31 denotes the part of the 2,950-bp *PvuII* fragment sequenced from the M13mp9 clones derived from the partially Bal 31-digested fragment.

**FIG. 4.** Structure of pTB9. The unfilled and filled double lines are pUB110 and pC194, respectively. The *PvuII* site and the *EcoRI* site in pUB110 are according to Jalanko et al. (15). Symbols for restriction sites are as described in the legend to Fig. 2.
other chemicals used were of purest commercially available quality.

**RESULTS**

Properties of strain BD294 carrying pJB74, pJB91, or pJB96. Some properties of BD294(pJB74), BD294(pJB91), and BD294(pJB96) are summarized in Table 1. Extracts from bacteria carrying pJB74 or pJB96 have increased levels of DNA polymerase activity, whereas only background levels are found in BD294(pJB91). The increased activity is not a gene dosage effect. The copy numbers of all three plasmids are similar and about half that of the vector pC194. The efficiency of plating (EOP) of SPO2 *susL244* is 1 on BD294(pJB74) and BD294(pJB96) compared with 0.1 for BD294(pJB91). After one cycle of growth of SPO2 *susL244* in the former strains, 10 to 20% of the progeny are wild-type recombinants, compared with 50 to 80% for BD294(pJB91).

This probably reflects the fact that in strains carrying pJB74 or pJB96, SPO2 *susL244* is complemented by DNA polymerase already present in the bacteria, whereas for pJB91 phage DNA replication can only occur after a functional gene *L* has been reconstituted by recombination between phage and plasmid. However, since BD294(pJB96) extracts contain high levels of DNA polymerase activity, gene *L* must be contained within the four HindIII fragments common to all three plasmids.

**Location of the *susL244* mutation.** HindIII fragments A and B were subcloned in pC194. The EOP of SPO2 *susL244* is ca. 20 times above the background level in BD294 carrying the B fragment (data not shown). Marker rescue of gene *L* from *EcoRI* fragments is known to be severely reduced compared with marker rescue of other genes (9). Part of gene *L* is thus located in HindIII fragment B.

**Subcloning of SPO2 gene *L* in PUB110.** The purpose of this

---

**FIG. 5.** Nucleotide sequence of gene *L* in pJB74. The nucleotide sequence of the anti-sense strand of pJB74 from the start of HindIII fragment E over the terminal *SusA* of pTB9 is shown. The positions of some pertinent restriction sites are indicated; the capital letter after HindIII denotes the start of the various SPO2 HindIII fragments (see Fig. 2). The inverted repeat in HindIII-E is denoted by arrows. EOP and rbs Pol indicate the proposed ribosome-binding sequences for the coding sequences for the proposed E Pol and the polymerase, respectively. Start E and Start Pol are the respective translational initiation triplets. Triplets marked with dots are the translational stop codons for the two open reading frames. One possible -35, -10 structure is indicated in HindIII fragment E by dashed lines.
TABLE 1. Some properties of strain BD294 carrying pJB74, pJB91, or pJB96

<table>
<thead>
<tr>
<th>Strain and plasmid</th>
<th>Relative DNA polymerase activity</th>
<th>Copy number</th>
<th>EOP of SPO2 susL244</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD294</td>
<td>1</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>BD294(pC194)</td>
<td>1</td>
<td>28, 39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BD294(pJB74)</td>
<td>2–3</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>BD294(pJB91)</td>
<td>1</td>
<td>10, 21</td>
<td>0.1</td>
</tr>
<tr>
<td>BD294(pJB96)</td>
<td>10–30</td>
<td>10, 11</td>
<td>1</td>
</tr>
</tbody>
</table>

* Plasmid copy numbers were determined by cesium chloride-ethidium bromide centrifugation of [3H]thymidine-labeled bacteria as described by Tanaka et al. (56).
* The EOP measured on different strains was related to the EOP measured on SR135, which carries the sup-3 suppressor gene.

The experiment was to determine the maximum size of gene L. pJB96 DNA was partially digested with Sau3A and ligated to BamHI-cleaved pUB110. The mixture was used to transform strain 3G18 selecting for kanamycin resistance. Among 1,000 transformants tested, 11 were found which pliated SP02 susL244 with an EOP of 1. Plasmids from these transformants were analyzed by agarose gel electrophoresis. The smallest plasmid observed, called pTB9, was transformed into BD294. Extracts of BD294(pTB9) have 10- to 20-fold increased levels of DNA polymerase activity. pTB9 contains SPO2 HindIII fragments B, F, and G and the proximal 429 bp of the 1,700-bp A fragment (Fig. 4). Adjacent to fragment B is a 481-bp pC194 fragment. We examined five other plasmids obtained by cloning partially digested pJB96 DNA into pUB110 and which express gene L. In addition to SPO2 DNA, they all contain the 481-bp pC194 fragment. pTB9 contains about 2,197 bp of SPO2 DNA, which represents coding capacity for a protein with a molecular size of 80 to 90 kd.

Nucleotide sequence of gene L. The above data suggest a simple model for expression of gene L in the three plasmids (see below). From this model we decided to sequence from pJB74 HindIII fragments E, B, G, and F, and the proximal 429 bp of fragment A. The strategy for sequencing is outlined above and is shown in Fig. 3. The sequence obtained is shown in Fig. 5. We identified two open reading frames. The first one starts at 356 bp in fragment E and proceeds 63 bp into fragment B. It has coding capacity for a hypothetical protein of ca. 10 kd, which we will call the E peptide. Six base-pairs upstream from the initiation codon ATG, there is a potential ribosome-binding site. It has a calculated free energy of interaction (ΔG) with the 3′ end of B. subtilis 16S RNA of ~16 kcal (~67 kJ)/mol. This value is similar to those reported for a number of other gram-positive ribosome-binding sites (23). The second open reading frame...
starts with TTG, which is a common initiation triplet in gram-positive bacteria (18). This TTG overlaps with the terminal codon and the termination triplet for the E peptide. Ten base pairs upstream from the TTG is a potential ribosome-binding site (ΔG = -15 kcal [−63 kJ]). This second open reading frame has coding capacity for a 72-kd protein (72486D). It extends through HindIII fragment B, across the fragments G and F, and 238 bp into fragment A. It terminates with three consecutive termination triplets: TAA TAG TAG. The Sau3A site used in the construction of pTB9 (Fig. 4) is located 181 bp downstream from the translational stop. Transcription of the E peptide open reading frame should start in HindIII fragment E as possibly also transcription of the 72-kd open reading frame. Between positions 27 and 48 in fragment E there is a guanine-cytosine rich inverted repeat followed by five consecutive pyrimidines. This structure is typical of some transcriptional stop signals (25, 27, 39). Several potential −35 and −10 promoter sequences can be found between the inverted repeat and the proposed ribosome-binding site for the E peptide transcript, e.g., a canonical −35 sequence (TTGACA) is found at positions 137 through 142, followed 16 bp downstream by a possible −10 sequence (AAGAAT) (27, 39). However, studies of transcripts from the cloned SP02 DNA are required to identify promoter sequences with any certainty (7, 23, 34).

DISCUSSION

B. subtilis BD294 carrying either pJB96 or its derivative pTB9, from which ca. 1,270 bp of HindIII fragment A have been deleted, are high SP02 DNA polymerase producers. From this fact, and the fact that a suppressible DNA polymerase mutation is located in HindIII fragment B, we conclude that the structural gene for SP02 DNA polymerase is contained with HindIII fragments B, G, and F and the proximal 429 bp of fragment A (A'; Fig. 4). Starting with a TTG in HindIII fragment B there is an open reading frame proceeding through fragments G and F and ending 238 bp into fragment A' with three consecutive translational termination triplets. It has the coding capacity for a 72-kd protein. All other open reading frames, reading in the same direction, have coding capacities for proteins of less than 10 kd. Also, none of those open reading frames is preceded by a reasonable ribosome-binding site. There is no open reading frame in the reversed order, starting in fragment A', which has coding capacity for a polypeptide larger than 18 kd. All three plasmids, pJB74, pJB91, and pJB96, carry a complete SP02 DNA polymerase gene, but the gene is not expressed in pJB91. The fact that all plasmids contain HindIII fragment A and that ca. 1,270 bp of fragment A, distal to HindIII fragment B, can be deleted in pJB96 without impairing expression of the gene excludes the possibility that the polymerase gene is read from a promoter located in fragment A. We conclude that the open reading frame coding for a 72-kd protein is the structural gene for SP02 DNA polymerase.

The present data suggest a simple model for transcription of the SP02 DNA polymerase gene, which accounts for the properties of the three plasmids studied here. Since pJB91 makes no polymerase although it contains a complete structural gene, transcription of the polymerase gene must be initiated from a promoter located outside HindIII fragment B, i.e., more than ca. 60 bp upstream from the initiation triplet TTG. In pJB74 the gene is transcribed from its natural promoter. We suggest that this promoter is located in HindIII fragment E. The presence of a possible transcriptional stop signal at positions 27 through 48 in fragment E (see Fig. 5) lends some speculative support to the idea that an SP02 transcript is initiated in fragment E.

In pJB96 the polymerase gene is removed from its natural promoter and instead the gene is transcribed from a promoter in the 482-bp HindIII-Sau3A fragment from the vector pCI94. Within this fragment there is an open reading frame coding for a hypothetical D protein (13); in pJB96 this is fused in phase with the terminal part of the proposed E peptide gene.

In pJB91, the polymerase gene is not transcribed. The amount of SP02 DNA polymerase activity found in BD294(pJB96) is 5 to 10 times greater than that found in BD294(pJB74); this might simply reflect differences in promoter strengths. However, the structure of pJB74 suggests the possibility of a more complex control of DNA polymerase in SP02. Starting at 336 bp in HindIII fragment E, there is an open reading frame, preceded by a typical gram-positive ribosome-binding site, coding for a proposed E peptide of ca. 10 kd. The reading frame for the E peptide overlaps with the proposed ribosome-binding site for the polymerase gene and with part of the initiation triplet TTG. If the E peptide and the polymerase genes are transcribed from different promoters, with the polymerase promoter located in the E peptide open reading frame, the amount of polymerase gene transcript could be modulated by E transcription. On the other hand, if both the E peptide and polymerase are transcribed from a polycistronic mRNA, the amount of polymerase produced could be controlled also at the translational level. We are presently trying to examine these possibilities.

Finally, Young and coworkers have established a correlated genetic and restriction map of phage SP02 (9). By combining their data with the results presented in this paper, the location, size, and direction of transcription of the phage polymerase gene can be determined on the SP02 genome (Fig. 6).

ACKNOWLEDGMENTS

We thank Sven-Ake Franzén and Tove Harde for expert technical assistance.
LITERATURE CITED


